Journal of Medical and Biological Engineering
https://doi.org/10.1007/s40846-018-0437-3

ORIGINAL ARTICLE

Mammographic Image Classification System via Active Learning

Yu Zhao[1] · Dong Chen[1] · Hongzhi Xie[2] · Shuyang Zhang[2] · Lixu Gu[1]

Received: 12 March 2018 / Accepted: 28 June 2018

Abstract
Training an accurate prediction model for mammographic image classification is usually necessary to require a large number of labeled images. However, the manually acquiring rich and reliable annotations is known to be tedious and time-consuming process, especially for medical image. The advances in machine learning yielded a branch of technique, termed active learning (AL), which has been proposed for solving the problem of the limited training samples and expensive labeling cost, and has resulted in highly successful applications in many pattern recognition tasks such as image processing and speech recognition. In this article, a comparison is provided among the mammographic image classification systems, relying on traditional supervised learning, un-supervised learning and AL, aiming to obtain a system with low labeling cost. The experiments based on digital database for screening mammography demonstrate that the AL is able to minimize the labeling cost of mammographic image without sacrificing the accuracy of final classification system. In addition, some specific characteristics of mammographic image: file information and spatial feature, which are not available to the traditional AL methods, have been found to further decrease the labeling cost. In conclusion, we suggest that the AL is a reasonable alternative to supervised learning for the researchers in the field of medical image classification with limited experimental conditions.

Keywords Image classification · Active learning · Mammography · Labeling cost

1 Introduction

Breast cancer is becoming more and more severe due to the environmental pollution and stressfulness of life, and thus it is considered as a major health problem worldwide [1]. Clinical data shows that the cure rate of breast cancer will increase from less than 40 to 90% if it is early detected, indicating that early diagnosis is crucial in reducing morbidity and mortality [2]. Mammography is one of the most widely used imaging techniques for early breast cancer detection. Radiologists can utilize these images to estimate the clinical pathological stages of the patients [3].

The application of classification system can assist doctors to conduct automatic and accurate analysis of mammographic images and distinguish the mass and normal breast tissue. Supervised learning, as a mature technique with stronger applicability and generality, has been extensively employed in this kind of image classification system through bridging the semantic gap between mammographic images and their diagnosis information [4]. The selection of classification model and the extraction of classification feature are two critical steps of this technique for achieving good performance [5], which is also the first problem that is necessary to take into account in this article: *the establishment of a high-performance classification system for mammographic image.*

Furthermore, the establishment of supervised machine learning based classification system usually require the acquisition of sufficient high-quality labels of each sample in hand. The manual annotations from radiologists is necessary and unavoidable. However, the label work of a large number of images is a tedious and time-consuming job, and even more so for mammographic images. The

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s40846-018-0437-3) contains supplementary material, which is available to authorized users.

✉ Hongzhi Xie
xiehongzhi@medmail.com.cn

✉ Lixu Gu
gulixu@sjtu.edu.cn

[1] Image Guided Surgery and Virtual Reality Lab, School of Biomedical Engineering, Shanghai Jiao Tong University, 800 Dong Chuang Road, Shanghai 200240, China

[2] Department of Cardiology, Peking Union Medical College Hospital, Beijing 100005, China

Published online: 20 July 2018

annotation work of mammographic image requires certain expertise and is often error-prone. The same image may be declared normal by one radiologist and suspicious by another. Besides, the annotation work of mammographic image is also more complex than natural image. It is generally known that each screening mammographic examination needs to take the head-to-foot (craniocaudal, CC) view and angled side-view (mediolateral oblique, MLO) images of the breast both [6, 7]. To ensure the correct label of each unlabeled mammographic image, the radiologists not only require to evaluate the characteristic manifestation of suspected mass regions in hand, but also can further confirm its label through the information from another image—the image reflects the same breast tissue but in different view. Such identification process has been validated its effectiveness by plenty of literature [8]. But in the meantime, the burden of radiologists for label work has been further increased. Hence, *the labeling cost minimization of classification system for mammographic image* becomes the other issue of this article.

The contribution of this study is around the handling of above two issues. First, to alleviate the first issue, we design and realize a mammographic image classification system using histogram of oriented gradient (HOG) feature and support vector machine (SVM). Second, on the basis of this system, this paper first introduces the active learning (AL) technique into the mammographic diagnosis domain to solve the problem of labeling cost minimization. Through intelligently choosing small valuable subsets from the entire dataset during the learning process, AL has the potential to develop accurate prediction models with fewer labeling operations from domain experts. (3) Furthermore, according to two specific characteristics of mammographic image viz., spatial feature and file information, the process of existing AL method is modified for further decreasing the labeling cost.

The paper is organized as follows. We briefly review conventional system for mammographic image classification and AL method in Sects. 2.1 and 2.2. In Sect. 3 we respectively introduce every main step of the proposed classification system that includes preprocessing and candidate image extraction (Sect. 3.2), the HOG based feature extraction method (Sect. 3.3), SVM classifier (Sect. 3.4) and the improved AL method, which is specialized for mammographic image (Sects. 3.5 and 3.6). Large numbers of experiments have been conducted to demonstrate that the improved mammographic image classification system can achieve high performance with very few labeled samples in Sect. 4. Finally, we discuss the comparison between existing system and proposed system in Sect. 5 and conclude in Sect. 6.

2 Related Work

2.1 The Establishment of a Classification System for Mammographic Image

The reasonable feature extraction methods and the appropriate classification model selection are two important factors to address the first issue of this paper. For all types of classification model, SVM is the most widely used in the mass and no-mass classification because of its good discriminative and forecasting ability. Most of the relevant literature about the classification system of mammographic images is dependent on the combination of SVM and various methods of feature extraction, which can be roughly divided into texture and shape feature. To the best of our knowledge, Lladó et al. [9] in 2009 first introduced SVM as the classification model for mammographic image classification, and their proposal is the use of local binary patterns (LBPs) for representing the textural properties of the masses, which assists their system to achieve 0.91 of AUC from 1792 images on digital database for screening mammography (DDSM). In 2011, with only 322 images in MIAS, Buciu and Gacsadi [10] used Gabor wavelets as the texture feature in their SVM based classification system and obtained 0.78 of AUC. In the same year, through the shape descriptors and the geostatistic functions as its shape and texture feature, the system in paper [11] was tested on the DDSM database with 3484 images and obtained sensitivity of 80% and AUC of 0.87. In 2013, Junior et al. [12] presented a method for mammographic image classification problem using texture features extracted as several diversity indices from images and SVM. As they reported, their method can reach 76–100% of accuracy with the experiments on 1600 images on DDSM. In 2015, the taxonomic diversity index and the taxonomic distinctness, which were originally used in ecology, were used as the texture feature descriptor of 3404 images in the de Oliveira et al.'s work [13], prompting their SVM based classification system achieves 98.33% of accuracy and 98.39% of sensitivity. The latest literature [14] in 2017, through the computation of binarized statistical image features and variants of LBP from the segmented images, obtained 97.12% sensitivity, 0.98 AUC on DDSM with 1781 images, and 95.12% sensitivity, 0.95 AUC on MIAS with 312 images.

Apart from SVM, there are still some other kind of classification model that can be brought into assist with the classification system of mammographic image, e.g., the linear discriminant analysis along with in efficient coding the work of Costa et al. [15] (reached 90.07% of accuracy from 5090 images on DDSM), the BP network along with GLCM in paper [16] (reached 98.8% of accuracy and

0.9945 of AUC from 2576 images on DDSM), the ANN network along with shape and texture features in paper [17] (reached 89.28% of accuracy and 0.928 of AUC from 330 images on MIAS), the C4.5 decision tree along with the shape feature extracted by the eigenfaces approach in paper [18] (reached 0.84 of AUC from 588 images on DDSM) and the C5.0 decision tree along with the shape feature vector that consists of 17 shape and margin properties [19] (reached 87.6% of accuracy from 224 images on DDSM). In the recent work proposed by Raghavendra et al. [20], after the feature extraction based on Gabor wavelet and the data reduction based on locality sensitive discriminant analysis (LSDA), the authors simply used most of popular classification model one by one, including Decision Tree, Linear Discriminant Analysis, Quadratic Discriminant Analysis, k-Nearest Neighbor, Naïve Bayes Classifier, Probabilistic Neural Network, SVM, AdaBoost and Fuzzy Sugeno, and selected the classifier with the highest performance as their final result (reached 98.69% of accuracy from 690 images on DDSM).

As the most special classification model, convolutional neural network (CNN) also has been used for mammographic image classification. Different from other classification models, the major advantage of CNN is its independence from prior knowledge and human effort in feature design, which means that the method of feature extraction is no longer needed. In fact, as early as in the mid-1990s, CNN has been used for the classification of mass and normal breast tissue in paper [9]. However, the performance (0.87 of AUC with 672 images) seems less than ideal under the limitation of conditions at that time. With the upgrading of hardware equipment, technological progress of CNN was obviously obtained in recent years. Kooi et al. [4] constructed a six-layer CNN network, and their learning features were subsequently extracted from the final layer of CNN and concatenated with the location and context features. Their system finally reaches 0.941 of AUC, after training on 45,000 images, which outperforms the other traditional system as they reported. In the same year, Jiang et al. [21] explored the technique of transfer learning to tackle the same problem. The CNN of GoogLeNet and AlexNet were utilized and reached 0.88 of AUC on a large-scale visual database in their work.

2.2 The Labeling Cost Minimization of Classification System for Mammographic Image

With respect to the second goal, few studies are conducted in the associated filed of labeling cost. From our point of view, there are two solutions can be considered to instead the traditional supervised learning way: the unsupervised learning algorithms and the AL algorithms.

The commonly used unsupervised learning algorithms include expectation–maximization algorithm [22], self-organizing map (SOM) [23] and adaptive resonance theory [24]. They are all trying to find hidden structure in dataset, which seem to be able to solve this issue once and for all, because they do not need pre-determined categorizations. Yet, the classification accuracies of unsupervised learning algorithms are usually not satisfactory owing to no prior knowledge required in the learning process, which is very likely to be against the first goal.

In contrast, AL are well tradeoff definition. During the learning process, AL algorithms can intelligently select small valuable subsets from the entire dataset for labeling, and thus has the potential to develop accurate prediction models with less labeling operations from domain experts [25, 26]. According to the different sample selection strategies (SQSs), eight typical AL algorithms are predominantly reported in the literature: (1) *diversity* [27]: collects the unlabeled samples that have greatly difference with the existing labeled samples, (2) *entropy* selects the highly informative samples using the entropy [28], (3) *TED* [29] a novel concept for AL, whose selected samples are most representative that can represent each data in a linear combination, (4) *Margin SVM* [30] calculates the distance from the samples to the separating hyperplane given by SVM, (5) *multiple view* (*MV*) chooses the high-quality samples by the controversy [31], (6) *query by committee* (*QBC*) filters informative queries from a random stream of inputs to select high-quality samples [32], (7) *uncertainty* [33] its selection strategy queries the samples whose posterior probability of being positive is nearest 0.5, (8) *expected model change* (*EMC*) [34] choose the samples, imparting the greatest change to the current model.

AL algorithms have been testified and applied in many domains including language processing [35], image processing [36], recommendation systems [37] and information retrieval [38]. However, to the best of our knowledge, AL methods have not yet been used in mammography classification system.

3 Materials and Methods

The entire process of mammographic image classification system with improved AL method can be described as Fig. 1.

3.1 Data Sources

The involved mammograms for testing are selected from DDSM [39, 40], a resource for use by the mammographic image analysis research community, were applied to establish and validate the AL algorithms embedded classification systems. The entire DDSM contains 2620 cases in 43

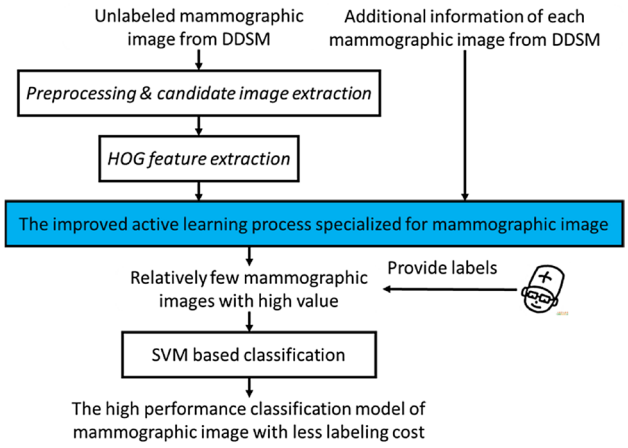

Fig. 1 The entire process of mammographic image classification system with improved AL

volumes, and each case collects four images for a single patient from a single exam including left CC, left MLO, right CC and right MLO.

3.2 Preprocessing and Candidate Image Extraction

The objective of this step is to greatly remove the unwanted elements in the mammographic images, thereby reducing the computational complexity and improving the performance of classification system. Since all mammograms on DDSM are compressed in. LJPEG format, which cannot be directly processed, they should be first decompressed and manipulated into TIFF format, through the approach outlined in the article [41]. Second, as shown in Fig. 2b, c, the foreground of each mammogram is separated from background by means of k-mean and morphological operation. Then, this foreground should be further enhanced through histogram equalization as the Fig. 2d. At last, through employing the pixel based Random Forest classifier with five kinds of pixel features [42] and seed point-based segmentation algorithm as proposed in literature [4], 3787 candidate images with various sizes can be extracted from these enhanced mammograms with background removed and construct an entire dataset T after the size of all of these candidate images have been unified as 128×128. Moreover, as the supplement for traditional method, the extraction position and mammogram source of each candidate image are also recorded as a list of image information, which will be used in the next step.

Here are some representative candidate images used for establishing the classification system in Fig. 3, where the first line is the benign and cancer masses, whose label equals -1. The second line is the normal tissue, whose label equals $+1$.

3.3 HOG Feature Extraction

The HOG is a feature descriptor proposed by Dalal and Triggs [43] used in computer vision and image processing for object detection. According to this paper, the main idea of HOG feature is to evaluate the well-normalized local histograms of image gradient orientations in a dense grid as the representations of its local area. The steps of HOG feature extraction can be described as Fig. 4: (1): first, the color space of input candidate images is normalized by the gamma color. (2) Second, the gradient of the image at each pixel is calculated for capturing the contour information (both its magnitude and direction). (3) Third, the entire image can be divided into several cell histograms, and the descriptor of each cell histogram can be obtained through the weighted vote for an orientation-based histogram channel on each pixel within this cell histogram. (4) Then, the cells are required to be grouped together into larger, spatially

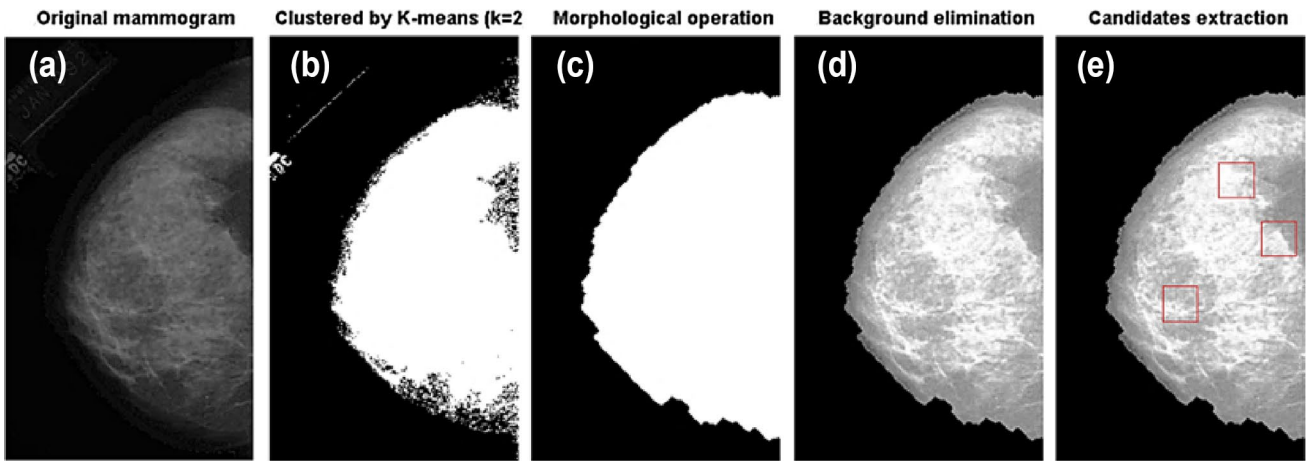

Fig. 2 The process of preprocessing and candidate image extraction

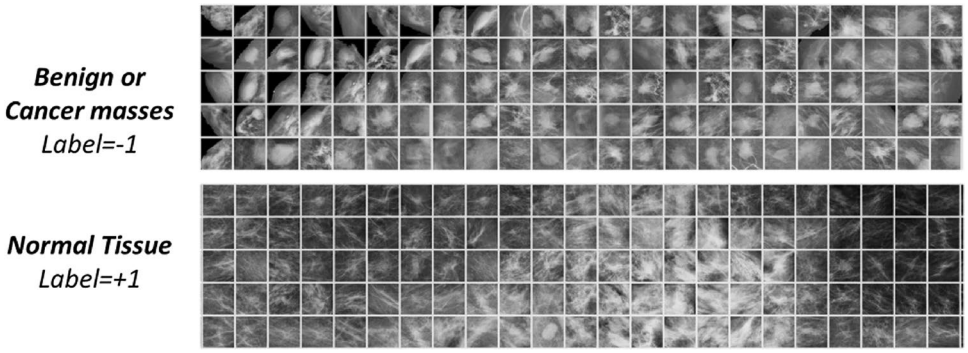

Fig. 3 Some representative candidate images

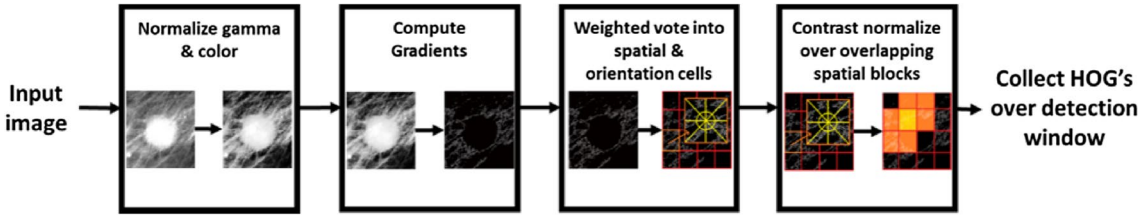

Fig. 4 The process of HOG feature extraction

connected blocks, and the contrast within each block also need to be normalized. (5) Finally, the HOG descriptor can be defined as the concatenated vector of the normalized cell histograms from all above blocks. Of particular note is that the dimension of HOG feature is up to 1800, so applying principal component analysis (PCA) [44] for dimension reduction (reserve 95% principal component) is necessary for removing redundant information, and also decreasing the running time of the training procedures. Some literatures [45, 46] also prove that the combination of SVM and PCA can achieve a better performance than using SVM only. To above collected candidate images, according to the recommended by paper [43, 47], the size of cell histograms and the size of block is respectively set to 8×8 pixel and 3×3 cell, the number of orientation bins is set to 8 and each weight for weighted vote for the orientation-based histogram channel is set to 1.

Compared with other common features used for mammographic image classification (e.g., shape and texture feature), HOG feature has two potential advantages: (1) as mentioned in [43], HOG feature can make a good description of local shape information, and is in better invariance to changes in transition, rotation, illumination and shadowing. Therefore, HOG feature might be the appropriate choice for the mammographic image with variform local structure that is prone to interference condition. (2) Most of common kinds of feature need the accurate image segmentation of tissue area before the feature extraction. However, the automatic segmentation of breast tissue in mammographic patches is not guaranteed to be reliable all the time, and the manual segmentation is even more inefficient than label work. If the radiologists had plenty of time for manual segmentation, the further label work of this image would not be hard. Then, the reduction of labeling cost would become meaningless. Conversely, as to the HOG feature, the accurate segmentation of tissue area is no longer necessary, which precisely coincides with the main theme of this paper.

3.4 The Classification Based on Support Vector Machine

SVM is introduced in this paper as the basic classification model in AL process, not only because it is the most common model used in mammographic image classification problem as mentioned before, but also because it indeed is the most appropriate for solving this kind of learning problem with small training set of samples. As the described in its earliest literature [48], SVM can efficiently perform a non-linear classification using what is called the kernel trick, implicitly mapping their inputs into high-dimensional feature spaces, and thus has higher efficiency and performance than other classification models.

3.5 Active Learning Algorithms for General Problem

The AL algorithms for general problem can be described as the following Fig. 5.

The input of conventional AL algorithms is an entire data set $\boldsymbol{T}$. In any iteration of AL process ($\boldsymbol{k}$th), this data set $\boldsymbol{T}$ can be divided into $\boldsymbol{U}^{k-1}$ and $\boldsymbol{A}^{k-1}$, which are respectively

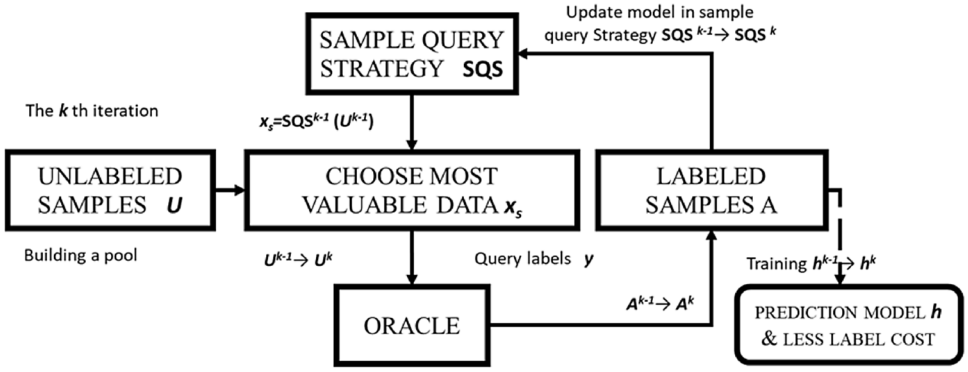

Fig. 5 Procedure of the AL algorithm for general problem

the existing unlabeled dataset and labeled dataset. $\boldsymbol{h}^{k-1}$ is the $\boldsymbol{k}-1$th prediction model that provided by $\boldsymbol{A}^{k-1}$. For any sample $\boldsymbol{x_i}$ in unlabeled dataset, its label $\boldsymbol{y}$ is unknown, which may be either -1 or 1. The main target of each iteration in AL process is to select one $\boldsymbol{x}_s$ with the highest value from $\boldsymbol{U}^{k-1}$ through one sample query strategy $\mathbf{SQS}^{k-1}$ and sent to the experts for labeling. Both the query strategy models and prediction model will be updated using the existing labeled samples. The iterations will not be suspended unless the termination criterion is satisfied. The termination criterion usually can be a performance goal or specified number of the labeled image. Then, the output of AL algorithms is just the collection of all selected samples from each iteration. Since the selected $\boldsymbol{x}_s$ in each iteration are all the most valuable one, only taking a few iterations, the classification model based on a few labeled sample selected by AL algorithms can achieve similar effects to the traditional supervised learning algorithms with plenty of labeled samples.

According to the different definition of high value, currently AL methods can be divided into following eight categories, and each of them has their special SQS. Here we list several common SQS in different AL methods used in this study, which defined as follows:

(1) *Diversity* [27]

$$x_s = SQS_{Diversity}\left(U^{k-1}\right) = \arg\min_{x_i \in U^{k-1}} \left(-1 * \arg\min_{x_j \in A^{k-1}} \frac{\kappa(x_i, x_j)}{\sqrt{\kappa(x_i, x_i)\kappa(x_j, x_j)}} \right), \tag{1}$$

where $\boldsymbol{\kappa}$ is the calculation of the Euclidean distance.

(2) *Entropy* [28]

$$x_s SQS_{Entropy}\left(U^{k-1}\right) = \arg\min_{x_i \in U^{k-1}} \left(-\sum_y P\left(y|x_i; h^{k-1}\right) \log P\left(y|x_i; h^{k-1}\right) \right). \tag{2}$$

(3) *Margin SVM* [30]

$$x_s = SQS_{M\arg in\_SVM}\left(U^{k-1}\right) = \arg\min_{x_i \in U^{k-1}} \left(w^T x_i + b\right)/||w||, \tag{3}$$

where $\boldsymbol{w}$ and $\boldsymbol{b}$ are the parameters from trained SVM $\boldsymbol{h}^{k-1}$.

(4) *QBC* [32]

$$x_s = SQS_{QBC}\left(U^{k-1}\right) = \arg\min_{x_i \in U^{k-1}} \left(\sum_y \frac{\sum_{j=1}^{C} I(h_j^{k-1}(x_i) = y)}{C} \log \frac{\sum_{j=1}^{C} I(h_j^{k-1}(x_i) = y)}{C} \right), \tag{4}$$

where $\boldsymbol{h}_j^{k-1}$ are $\boldsymbol{C}$ different kinds of prediction models with competing hypotheses in $\boldsymbol{k}-1$ iteration, and $\mathbf{I(.)}$ is an indicator function that is equal to one if conditions within the parentheses are satisfied.

(5) *MV* [31]

$$x_s = SQS_{MV}\left(U^{k-1}\right) = \arg\min_{x_i \in U^{k-1}} \left(\sum_y \frac{\sum_{j=1}^{C} I(h^{k-1}(D_j(x_i)) = y)}{C} \log \frac{\sum_{j=1}^{C} I(h^{k-1}(D_j(x_i)) = y)}{C} \right), \tag{5}$$

where $\boldsymbol{h}^{k-1}$ is a particular prediction model in $\boldsymbol{k}-1$ iteration, and D_j are $\boldsymbol{C}$ different feature combinations of sample x_i.

(6) *Uncertainty* [33]

$$x_s = SQS_{Un}\left(U^{k-1}\right) = \arg\min_{x_i \in U^{k-1}} P\left(\left(\arg\min_{y} P\left(y|x_i\right); h^{k-1}\right)|x_i; h^{k-1}\right). \quad (6)$$

(7) *EMC* [34]

$$x_s = SQS_{EMC}\left(U^{k-1}\right) = \arg\min_{x_i \in U^{k-1}} \sum_{y} P\left(y|x_i; h^{k-1}\right) \nabla l\left(A^{k-1} \cup x_i,\ y; h^{k-1}\right), \quad (7)$$

where ∇l is the gradient of the objective function l.

Beside the above seven **SQS**, new research reports a new kind of AL algorithm and it derivations, which suggest that the data points, which can capture the intrinsic information of the whole data collection, is the most valuable, termed 'transductive experimental design'. The main difference between it and above seven methods is the transductive experimental design-based AL methods can directly obtain any number valuable samples without iteration, and for the entire unlabeled dataset $\boldsymbol{U^0}$, the part of samples $\boldsymbol{A^1}$, which contains $\boldsymbol{m}$ valuable samples that need to be labeled, can be expressed as below.

(8) *TED* [29]

$$A^1 = SQS_{Ted}(U) = \arg\min_{A^1 \subset U^0} \sum_{i=1}^{|U^0|} \left(\left\|x_i - A^1 b_i\right\|_2^2 + \lambda \left\|b_i\right\|_2^2\right), \quad s.t.\ \left[b_1, \ldots, b_{|U^0|}\right] \in R^{m \times |U^0|} \quad (8)$$

where $|.|$ means the length of vector, $|A^1| = m$, and $\|.\|_2$ is the l2-norm of the vector.

More detailed information of these AL algorithms can refer to the literature [27–34]. In general, each query strategy represents the different definitions of the measure of 'valuable', leading to different selection of samples. Since the optimal sample selection not only lies on the algorithm, but also is closely related to the sample distribution, the best AL algorithm for mammographic image classification system cannot be asserted, and still needs to be confirmed through the experiments as the next section.

3.6 Active Learning Algorithms for Mammographic Image Classification

With additional studies of mammographic image, we also found that the labeling cost can be further reduced if we leverage the characteristics of mammographic image. Unlike nature images, there may be two kinds of interrelation between two candidate mammographic images, and two candidate images in one kind of above interrelation represent the same part of breast tissue in the mammogram, and their label must be the same.

The first interrelation is caused by the pixel-based candidate image extraction method mentioned above. Two candidate images in this interrelation are extracted from the same mammogram and same view but slightly different extraction positions, and there is plenty of overlap in these two images as the blue and red box in Fig. 6. Besides, the second interrelation between two candidate images can be regarded as one of the unique characteristics of mammographic image, which has been discussed in many literatures [6]. Although the candidate images in the second interrelation are extracted from two different source mammograms, they are actually same part of breast tissue in different view

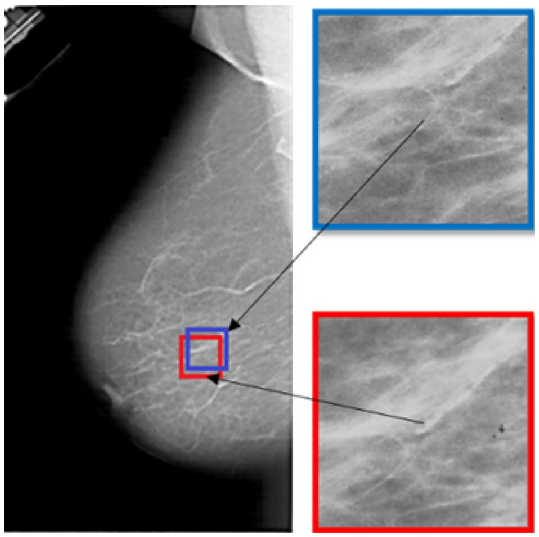

Fig. 6 The first interrelation

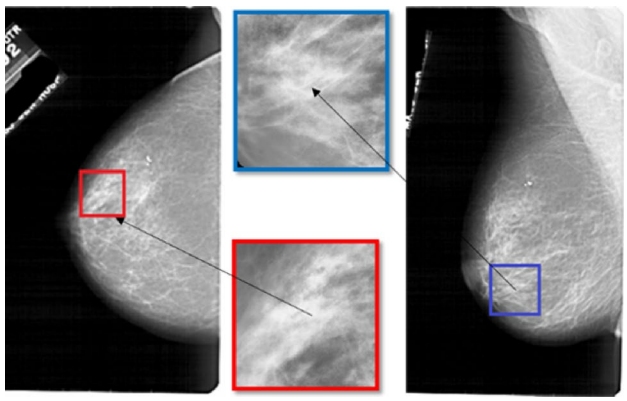

Fig. 7 The second interrelation between two candidate images between two candidate images

as the blue and red box in Fig. 7. The only thing we know is that their source mammograms must have the same serial number but different view (MLO and CC), and we define such a pair of mammograms as $\boldsymbol{M}^*$ and $\boldsymbol{M}$ here.

Inevitably, the candidate images with one kind of above relation will share the same label. It is not necessary to label these two candidate images twice, and thus can significantly decrease the labeling cost. Then, the above issues are turned into a problem: how to determine if there is the first or second interrelation between the currently selected image $\boldsymbol{x}_s$ and each labeled candidate image $\boldsymbol{x}_j \in \boldsymbol{A}^{k-1}$.

The spatial information can be a good indicator of whether first interrelation exists between any two candidate images. That is to say, for each candidate image $\boldsymbol{x}$, we need to record its additional spatial information including that the location of its central point in source mammogram (***x.Mp***), the view of its source mammogram (***x.View***) and the serial number of its source mammogram (***x.Ms***) in the step of candidate image extraction. In any iteration of AL process, the selected image $\boldsymbol{x}_s$ can be neglected without labeling if there is an image $\boldsymbol{x}_j$ existing in $\boldsymbol{A}^{k-1}$, which has the *first interrelation* with $\boldsymbol{x}_s$ as set of formula (9):

$$\begin{cases} x_s.Ms = x_j.Ms, \\ x_s.View = x_j.View, \\ |x_s.Mp - x_j.Mp| \le \varepsilon, \end{cases} \tag{9}$$

where ε can be set as the half length of the candidate image diagonal (ε = 90 in this article), and the |.| is the Euclidean distance formula.

Comparing with the *first interrelation*, the *second interrelation* is more difficult to realized. The prerequisite of the *second interrelation* is to find a solution that is able to accurately search image $\boldsymbol{x}^*$ from mammogram $\boldsymbol{M}^*$ through the image $\boldsymbol{x}$ *in* mammogram $\boldsymbol{M}$. In consideration that there is no perfect automated method to achieve this solution, we still suggest that this step should be done by radiologists. Then, subsequent work could be addressed as the solution of the problem of first relation, and the *second interrelation* can be defined as set of formula (10). As the label work of each candidate images is supposed to compare both the two views of mammogram, the work burden of radiologists will not increase.

$$\begin{cases} x_s.Ms = x_j.Ms, \\ x_s.View \neq x_j.View, \\ |x_s^*.Mp - x_j.Mp| \le \varepsilon. \end{cases} \tag{10}$$

Then, without considering repeated iterations as shown in Fig. 5, the process of the *k*th iteration in our improved AL method can be described as Fig. 8, whose inputs are the labeled dataset $\boldsymbol{A}^{k-1}$ and unlabeled dataset $\boldsymbol{U}^{k-1}$ with their corresponding spatial information obtained from the

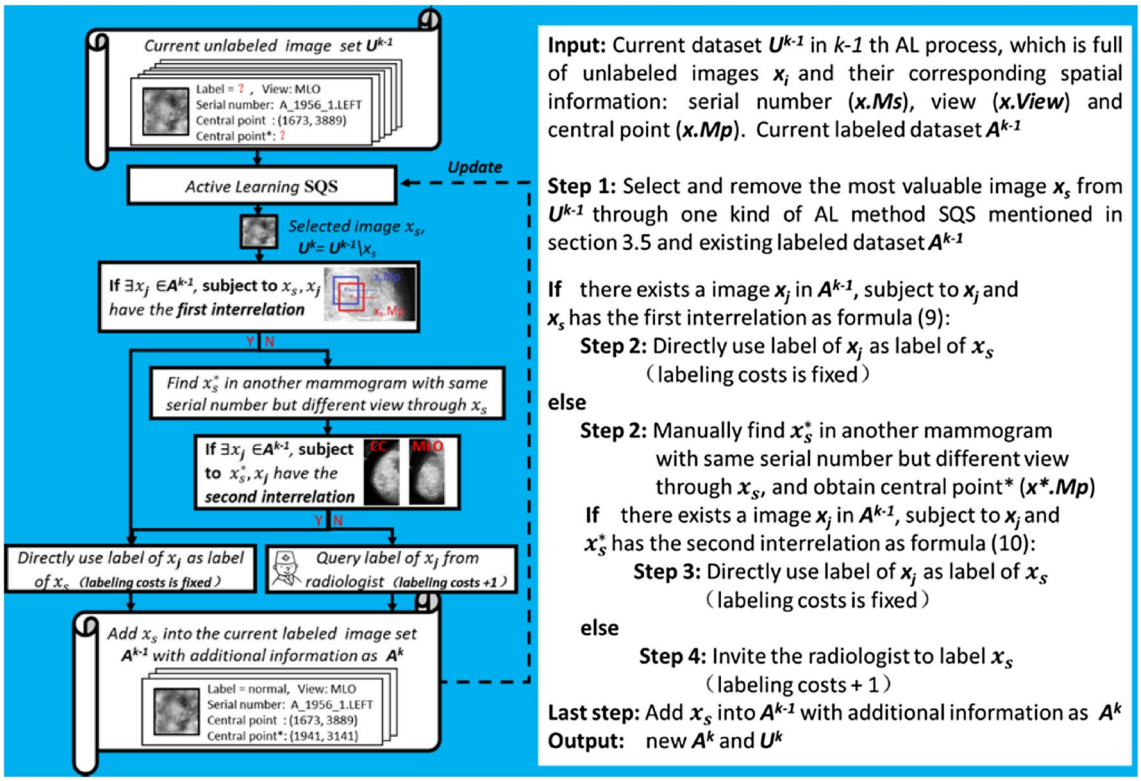

Fig. 8 One iteration of our improved AL method specially designed for mammographic image

previous iteration. The outputs are the new $\boldsymbol{A}^k$ and $\boldsymbol{U}^k$, which will be used as the input in the next iteration.

4 Experiment

4.1 Experimental Environment

All operations in this study were executed in Matlab R2014a software (The Mathworks, Inc., Natick, MA, USA), which was installed in the PC with the Intel Core i3-2100 CPU (3.10 GHz) and 3 GB memory. LIBSVM supported by http://www.csie.ntu.edu.tw/~cjlin/libsvm/ was applied to train the SVM classification models with radial basis function kernels for all classification tests [49], and the penalty parameter and gamma parameter are respectively fixed to 1 and 0.125.

To ensure the performance of each method and their operating reliability, all involved methods are repeated for 10 times, and their averages and standard deviation are calculated for analysis. In each time, all 3787 candidate images mentioned in Sect. 3.2 will be randomly divided into a training set with 50% of the samples and a test set with the rest 50% of the samples.

4.2 Validation of the Methods

Moreover, four metrics namely accuracy, precision, recall rate, area under receiver operator characteristic (ROC) curve (AUC) and labeling cost are used to estimate the effectiveness of these classification models. The first three can be calculated using the following equations. As mentioned in paper [50], the AUC value is equivalent to the probability that a randomly chosen positive example is ranked higher than a randomly chosen negative example, which also can be defined as the area under the ROC curve. The ROC curve describes the ability of the classification model to correctly differentiate the set of images into two classes based on the true-positive fraction (sensitivity) and false-positive fraction (1 − specificity). Due to the AL process involves plenty of classification models, the AUC value is more appropriate than ROC curve for the evaluation of AL algorithms.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}, \tag{11}$$

$$\text{Precision} = \frac{TP}{TP + FP}, \tag{12}$$

$$\text{Recall} = \frac{TP}{TP + FN}, \tag{13}$$

where TP, TN, FP, FN are representative of true positives (the sample is positive and the predicted class is positive), true negatives (the negative sample is classified as negative), false positives (the negative sample is classified as positive), and false negatives (the sample is positive and the predicted class is negative), respectively. A well approaches for mammographic image classification can reach high accuracy, precision, recall rate and AUC with less labeling cost.

4.3 Experimental Process

In the Experiment A, the comparison of experiments was carried out among eight common AL algorithms and random selection mentioned before in Fig. 9, including: *Diversity*, *Entropy*, *TED*, *Margin SVM*, *MV*, *QBC*, *Uncertainty* and *EMC*. Before the AL process, we randomly select two candidate images of normal tissue and two candidate images of mass as the first batch of selection, and in each iteration of AL process, one most valuable image will be selected for querying label. The x-coordinate in the figure below is the labeling cost, and the y-coordinate is the performance of each method. Moreover, we also record the CPU time of each method in Table 1.

In the Experiment B, we make a contrast test on the improved AL methods with the original ways in Fig. 10. The former is denoted as solid lines, and the later are dotted lines. The involved AL methods in this part are the best three selected in the Experiment A.

In Experiment C, we compare the best improved AL algorithm among the eight algorithms to the unsupervised and supervised approaches, which are the SOMs (iteration = 50) and SVM, respectively. In every repeated test of both the control methods and the improved AL method, 3787 candidate images have been randomly divided into a training set with 1893 training images and a test set with the rest 1894 samples. The procedure of comparison is presented in Fig. 11, and the mean and standard deviation of experimental results will be recorded in the Table 2, where 'Labeling cost' means the number of unlabeled images has been labeled before the establishment of its corresponding classification model.

4.4 Experimental Results and Discussion

Figure 9 shows the performance metrics of mammographic image classifiers modeled by AL algorithms and random selection. It indicates that the results of most of AL algorithms are superior to the random selection except *MV*, it's probably because the involved feature combinations we designed in this AL algorithm are all randomized, which are far from optimal. Moreover, it also can be observed that in the different stages of AL process, the AL algorithms with the best performance are also different: *Diversity* in the early stages (labeling cost 5–75), *Margin SVM* in the

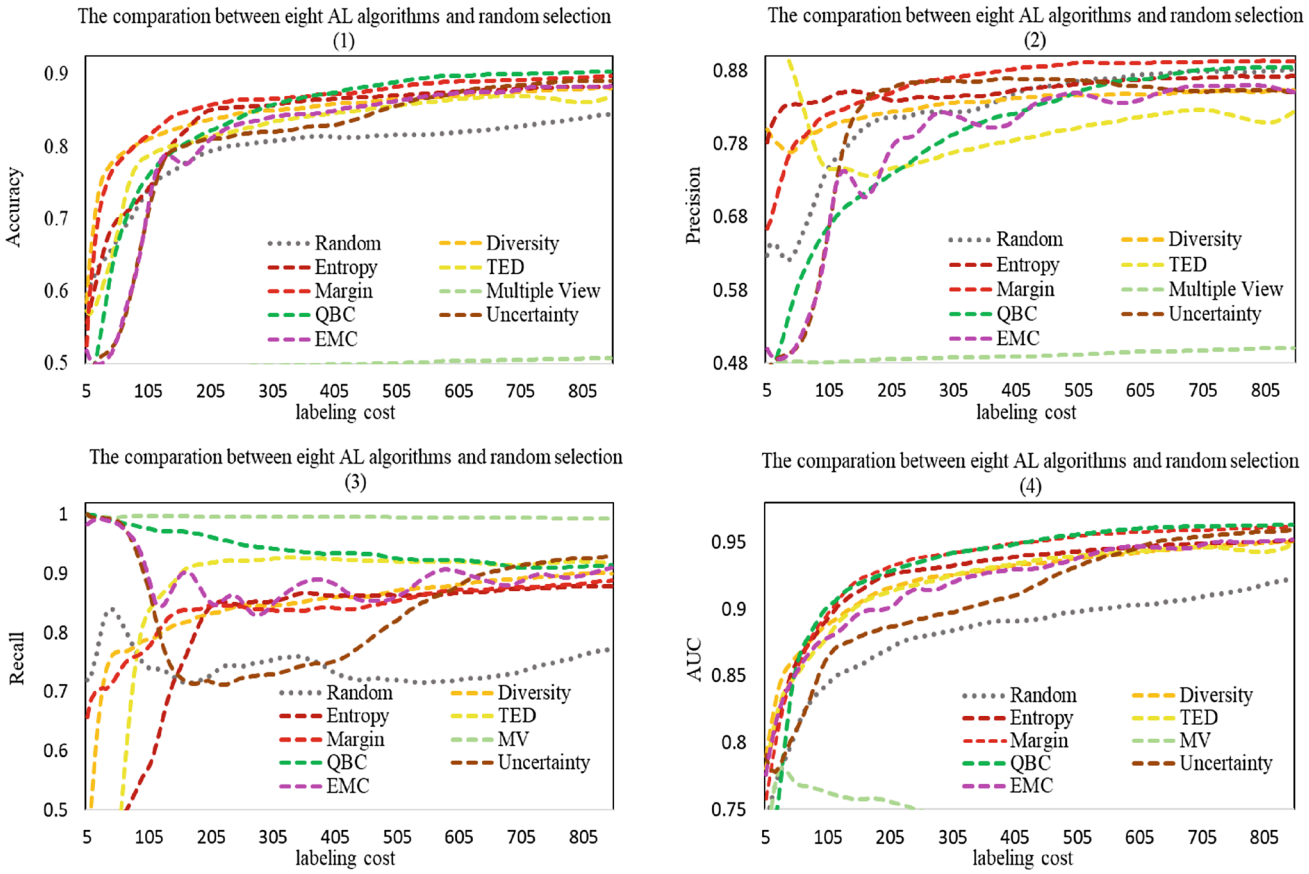

Fig. 9 The comparison between eight AL algorithms and random selection

Table 1 Comparing the CPU time of each methods

Random	Diversity	Entropy	TED	Margin SVM	MV	QBC	Uncertainty	EMC
19.0951	328.24	102.94	1766	40.6	49.35	373	102.6	63

middle stages (labeling cost 80–400), and *QBC* in the later stages (labeling cost 405–850). That exactly matches for that description in the recent literature [51] that one AL algorithm can only guarantee its optimal performance over a period of time in the entire AL process, and the optimal period differs for each algorithm.

Table 1 further indicates that the *Margin SVM* takes the shortest time among nine methods. Conversely, *TED*, *Diversity*, which need iterate over entire unlabeled dataset and *QBC*, which involves multiple classification models, are much slower. The CPU time of these methods cost several times more than *Margin SVM*. From the above results, taking into consideration of both the time and performance, we can conclude that *Margin SVM* is the optimal AL algorithm for mammographic image classification.

Figure 10 demonstrates that with the modifications as mentioned in Sect. 3.6, the improved *QBC*, *Diversity*, and *Margin SVM* really offer advantages over their original version, although it's not a big advantage. To our view, it is probably because the opportunities that the selected candidate image satisfies the formulas (9) and (10) are not many.

In Fig. 11, the AL algorithm (improved *Margin SVM*) demonstrates the more excellent performance than the SOM with the accuracy, precision and recall rate. With the increasing the labeling cost, its advantage is becoming more and more obvious. In addition, AL can achieve the similar performance of supervised learning with less than half of labeling cost (almost 44%) in mammographic image classification problem.

In this paper, we also made a series of comparisons between proposed AL based mammographic image classification system and the existing methods in the recent literatures in Table 3. A special performance index, termed 'active-learning labeling cost' (ALC) was created in this article. The ALC of one paper means the minimum labeling cost that the proposed AL based system needs, for achieving the better or

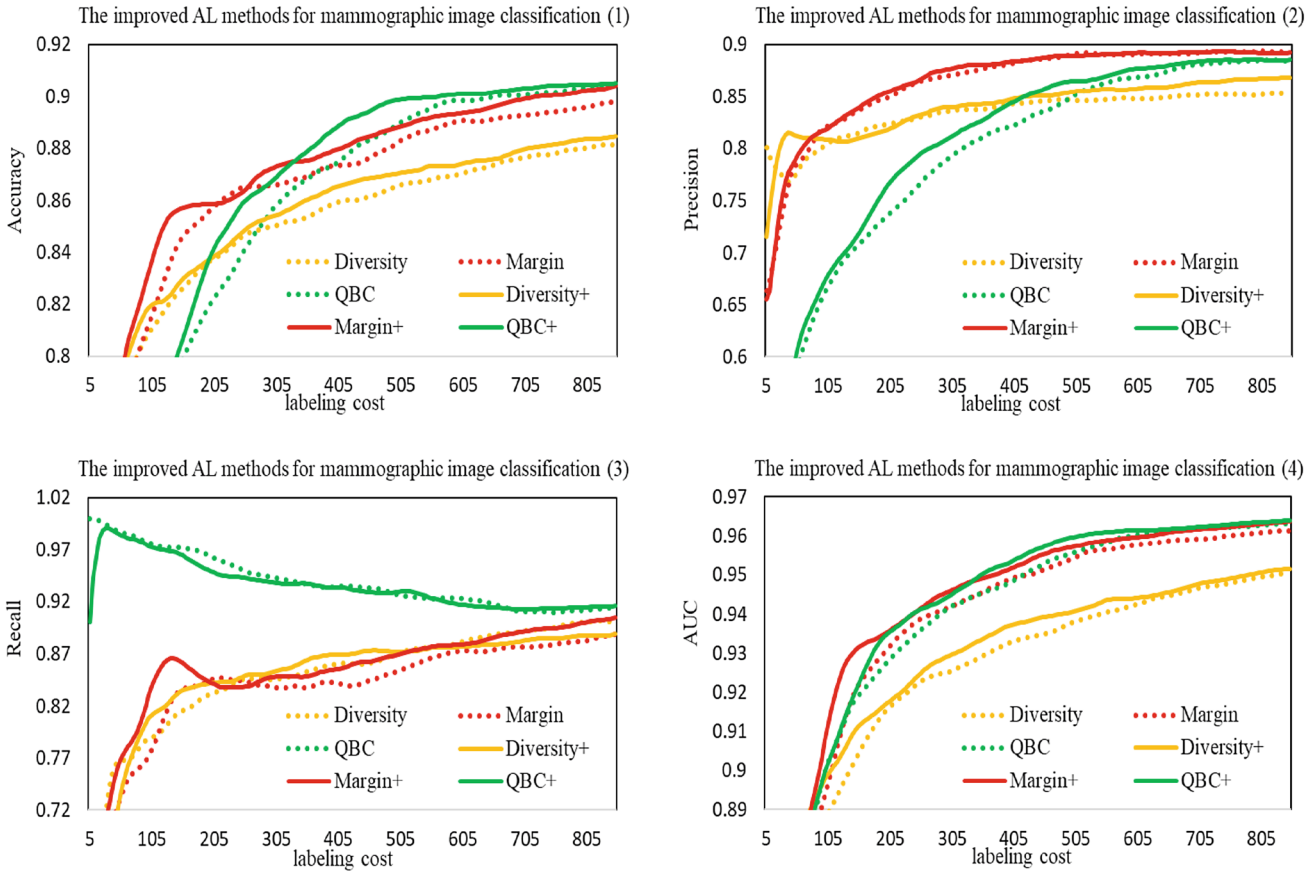

Fig. 10 The improved AL methods for mammographic image classification

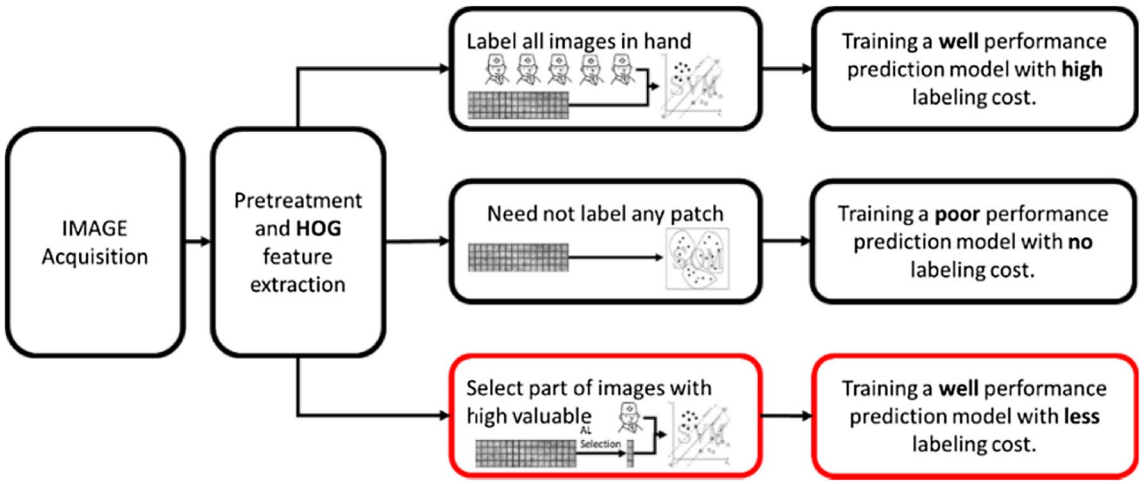

Fig. 11 The basic process of mammographic image classification system based on unsupervised learning, supervised learning and active learning algorithms

equal effect than the performance records in this paper for the first time. From the Table 3, we can observe that the performance of proposed system is better than the most of research groups before 2014. However, in the next few papers, e.g., [13, 14, 16, 20], faced their almost perfect performance, our AL based mammographic image classification system appear little powerless. We consider the reasons are as follows: in order not to violate the intention of the AL algorithm, our system typically uses HOG feature to avoid the accurate image segmentation of breast tissue, but it also limit the selection of

Table 2 Comparing the mammographic image classification system based on unsupervised learning, supervised learning and active learning algorithms

	Unsupervised	Active learning					Supervised
Accuracy	62.45 ± 0.4100	66.77 ± 0.1341	77.64 ± 0.0333	83.98 ± 0.0073	88.78 ± 0.0019	90.5 ± 0.0025	90.55 ± 0.0025
Precision	64.84 ± 1.1300	58.81 ± 0.0956	68.51 ± 0.0378	76.43 ± 0.0118	84.3 ± 0.009	88.51 ± 0.0046	88.20 ± 0.0056
Recall	77.46 ± 1.5400	98.47 ± 0.0382	96.78 ± 0.0207	95.07 ± 0.0255	93.42 ± 0.0127	91.56 ± 0.0085	92.14 ± 0.0084
AUC	0.830 ± 0.0100	85.74 ± 0.0556	90.16 ± 0.0112	93.57 ± 0.0048	95.35 ± 0.0024	96.41 ± 0.0008	96.53 ± 0.0011
Labeling cost	0	50	100	200	400	850	1893

Table 3 The comparison to the existing methods of mammographic image classification system

Paper	Year	Feature and model	Performance	Label-cost	Dataset	ALC
[9]	1996	CNN architecture and texture feature	AUC = 0.87	672	Unknown	55
[18]	2006	C4.5 decision tree, KNN and shape feature	AUC = 0.84	588	DDSM	35
[9]	2009	SVM and LBP	AUC = 0.91 ± 0.04	1792	DDSM	125
[15]	2011	LDA and efficient coding	Accuracy = 90.07%	5090	DDSM	590
[10]	2011	SVM and Gabor wavelets	AUC = 0.78	322	MIAS	20
[11]	2011	SVM and shape and texture descriptors	AUC = 0.87	3484	DDSM	55
[12]	2013	SVM and diversity indexes spatial decompositions	Accuracy = 88.25%	1600	DDSM	375
[17]	2013	Neural networks and shape, density features	Accuracy = 89.28%, AUC = 0.928	330	MIAS	445 or 175
[19]	2013	C5.0 decision tree and shape and margin	Accuracy = 87.6%	224	DDSM	340
[16]	2014	BPNN and GLCM, discrete wavelet transforms	Accuracy = 98.8%, AUC = 0.9945	2576	DDSM	Can't
[13]	2015	SVM and taxonomic diversity, distinctness	Accuracy = 98.33%	3404	DDSM	Can't
[20]	2016	SVM and Gabor, LSDA	Accuracy = 98.69%	690	DDSM	Can't
[4]	2017	CNN and location, context	AUC = 0.94 + 0.02	45,000	Unknown	240
[21]	2017	GoogLeNet	AUC = 0.88	Unknown	Unknown	70
[14]	2017	SVM and BSIF, LBP	AUC = 0.98	1312	DDSM	Can't
Proposed		SVM, AL and HOG	Accuracy = 90.5, AUC = 96.41	850	DDSM	~

feature. Many more kinds of texture and shape feature with higher distinguish ability can't be used in our system. On the other side, the major goal of AL algorithms is to reduce the labeling cost rather than more accurate classification and thus don't improve performance. Of course, it also can be observed that our proposed mammographic image classification system can achieve sharp reductions in the labeling cost, as long as the performance can be achieved though HOG feature. Therefore, in our view, the proposed system with above performance is perfectly acceptable.

5 Discussion

The possible application perspectives of the establishment of the AL based mammographic image classification system can refer to well medical applications. According to the experimental results in above section, only with 44% of the original labeling cost (850/1893), the performance of our improved system is close to the system with conventional supervised machine learning model which is undoubtedly alleviates the burden of annotation work. Time and manpower costs are saved.

The experiments we designed in the above section not only demonstrated the viability of introduction AL algorithms in the training process of mammographic image classification system, but also indicated that the AL algorithms designed for general problems won't consider the specific characteristics of mammographic image and make their results suboptimal. In this study, on the basis of traditional AL process, the selected candidate images will be further screened through their spatial feature and file information. Only the most valuable and unconjecturable images will be submitted to radiologists for querying labels, and the labeling cost can be further reduced. It is worth mentioning that unlike natural images, spatial feature and file information are always available. Therefore, these reinforced versions of AL methods are also suitable for the other specific purpose in medical domain.

6 Conclusions

In this study, an efficient AL algorithm based mammographic image classification system was proposed for differentiating the mass and no mass images from mammograms. The classification system combining AL algorithm is proved to be applicable for breast mammographic image, and among all involved AL algorithms, Diversity, Margin SVM and QBC respectively has the best performance in the early, middle and later stages. Through the further improvement of these well performed AL algorithms based on the characteristics of mammographic image, the labeling cost of candidate images can be further decreased, and the overall performance is still outstanding (with only 850 images, its performance can achieve 90.5% accuracy, 88.51% precision, 91.56% recall and 96.41 AUC).

Nonetheless, our future studies will be launched from two aspects: first, for further improving the classification performance, two classification models will be established respectively for the candidate images in CC and MLO view, and the entire AL process also makes corresponding adjustments with the introduction of two-view models, second, we will also attempt to develop our own AL algorithm, which can always get the best performance, whether at the early, middle or later stages of AL process.

Acknowledgements This research is partially supported by the National Key Research Program of China (2016YFC0106200), the 863 national research fund (2015AA043203) of China, the National Natural Science Foundation of China (81301283, 61190124 and 61271318), and the special funding of capital health research and development with No. 2016-1-4011. The authors are grateful to the Massachusetts General Hospital, the University of South Florida, and Sandia National Laboratories, which provides DDSM as a resource for our experimental data. We also express our sincere gratitude towards Department of Computer Science in University of North Carolina at Charlotte for their free tech support.

Compliance with Ethical Standards

Conflict of interests The authors declare that there is no conflict of interests regarding the publication of this paper.

Ethical Approval The study doesn't involve human or animal subjects.

References

1. Oliver, A., Freixenet, J., Marti, J., Perez, E., Pont, J., Denton, E. R., et al. (2010). A review of automatic mass detection and segmentation in mammographic images. *Medical Image Analysis, 14,* 87–110. https://doi.org/10.1016/j.media.2009.12.005.
2. World Health Organization. (2012). *International Agency for Research on Cancer GLOBOCAN 2012: Estimated cancer incidence, mortality and prevalence worldwide in 2012*. Geneva: WHO.
3. Sickles, E. A. (1989). Breast masses: Mammographic evaluation. *Radiology, 173,* 297–303. https://doi.org/10.1148/radiology.173.2.2678242.
4. Kooi, T., Litjens, G., van Ginneken, B., Gubern-Mérida, A., Sánchez, C. I., Mann, R., et al. (2017). Large scale deep learning for computer aided detection of mammographic lesions. *Medical Image Analysis, 35,* 303–312. https://doi.org/10.1016/j.media.2016.07.007.
5. de Lima, S. M., da Silva-Filho, A. G., & dos Santos, W. P. (2016). Detection and classification of masses in mammographic images in a multi-kernel approach. *Computer Methods and Programs in Biomedicine, 134,* 11–29. https://doi.org/10.1016/j.cmpb.2016.04.029.
6. Bekker, A. J., Shalhon, M., Greenspan, H., & Goldberger, J. (2015). Learning to combine decisions from multiple mammography views. In *2015 IEEE 12th international symposium on biomedical imaging (ISBI)* (pp. 97–100). IEEE. https://doi.org/10.1109/isbi.2015.7163825.
7. Andersson, I., Hildell, J., Muhlow, A., & Pettersson, H. (1978). Number of projections in mammography: Influence on detection of breast disease. *American Journal of Roentgenology, 130,* 349–351. https://doi.org/10.2214/ajr.130.2.349.
8. Sickles, E., Weber, W., Galvin, H., Ominsky, S., & Sollitto, R. (1986). Baseline screening mammography: One vs two views per breast. *American Journal of Roentgenology, 147,* 1149–1153. https://doi.org/10.2214/ajr.147.6.1149.
9. Lladó, X., Oliver, A., Freixenet, J., Martí, R., & Martí, J. (2009). A textural approach for mass false positive reduction in mammography. *Computerized Medical Imaging and Graphics, 33,* 415–422. https://doi.org/10.1016/j.compmedimag.2009.03.007.
10. Buciu, I., & Gacsadi, A. (2011). Directional features for automatic tumor classification of mammogram images. *Biomedical Signal Processing and Control, 6,* 370–378. https://doi.org/10.1016/j.bspc.2010.10.003.
11. Sampaio, W. B., Diniz, E. M., Silva, A. C., De Paiva, A. C., & Gattass, M. (2011). Detection of masses in mammogram images using CNN, geostatistic functions and SVM. *Computers in Biology and Medicine, 41,* 653–664. https://doi.org/10.1016/j.compbiomed.2011.05.017.
12. Junior, G. B., da Rocha, S. V., Gattass, M., Silva, A. C., & de Paiva, A. C. (2013). A mass classification using spatial diversity approaches in mammography images for false positive reduction. *Expert Systems with Applications, 40,* 7534–7543. https://doi.org/10.1016/j.eswa.2013.07.034.
13. de Oliveira, F. S. S., de Carvalho Filho, A. O., Silva, A. C., de Paiva, A. C., & Gattass, M. (2015). Classification of breast regions as mass and non-mass based on digital mammograms using taxonomic indexes and SVM. *Computers in Biology and Medicine, 57,* 42–53. https://doi.org/10.1016/j.compbiomed.2014.11.016.
14. Kashyap, K. L., Bajpai, M. K., & Khanna, P. (2017). Globally supported radial basis function based collocation method for evolution of level set in mass segmentation using mammograms. *Computers in Biology and Medicine, 87,* 22–37. https://doi.org/10.1016/j.compbiomed.2017.05.015.
15. Costa, D. D., Campos, L. F., & Barros, A. K. (2011). Classification of breast tissue in mammograms using efficient coding. *Biomedical Engineering Online, 10,* 55. https://doi.org/10.1186/1475-925x-10-55.
16. Beura, S., Majhi, B., & Dash, R. (2015). Mammogram classification using two dimensional discrete wavelet transform and gray-level co-occurrence matrix for detection of breast cancer. *Neurocomputing, 154,* 1–14. https://doi.org/10.1016/j.neucom.2014.12.032.
17. Saki, F., Tahmasbi, A., Soltanian-Zadeh, H., & Shokouhi, S. B. (2013). Fast opposite weight learning rules with application in

breast cancer diagnosis. *Computers in Biology and Medicine, 43,* 32–41. https://doi.org/10.1016/j.compbiomed.2012.10.006.
18. Oliver, A., Marti, J., Marti, R., Bosch, A., & Freixenet, J. (2006). A new approach to the classification of mammographic masses and normal breast tissue. In *18th International conference on pattern recognition, 2006. ICPR 2006* (pp. 707–710). IEEE. https://doi.org/10.1109/icpr.2006.113.
19. Vadivel, A., & Surendiran, B. (2013). A fuzzy rule-based approach for characterization of mammogram masses into BI-RADS shape categories. *Computers in Biology and Medicine, 43,* 259–267. https://doi.org/10.1016/j.compbiomed.2013.01.004.
20. Raghavendra, U., Acharya, U. R., Fujita, H., Gudigar, A., Tan, J. H., & Chokkadi, S. (2016). Application of Gabor wavelet and Locality Sensitive Discriminant Analysis for automated identification of breast cancer using digitized mammogram images. *Applied Soft Computing, 46,* 151–161. https://doi.org/10.1016/j.asoc.2016.04.036.
21. Jiang, F., Liu, H., Yu, S., & Xie, Y. (2017). Breast mass lesion classification in mammograms by transfer learning. In *Proceedings of the 5th international conference on bioinformatics and computational biology, 2017* (pp. 59–62). ACM. https://doi.org/10.1145/3035012.3035022.
22. Do, C. B., & Batzoglou, S. (2008). What is the expectation maximization algorithm? *Nature Biotechnology, 26,* 897. https://doi.org/10.1038/nbt1406.
23. Kohonen, T. (1998). The self-organizing map. *Neurocomputing, 21,* 1–6. https://doi.org/10.1016/s0925-2312(98)00030-7.
24. Braspenning, P. J., & Thuijsman, F. (1995). *Artificial neural networks: An introduction to ANN theory and practice* (Vol. 931, pp. 101–117). Berlin: Springer.
25. Settles, B. (2010). *Active learning literature survey* 52-11. Madison, WI: University of Wisconsin.
26. Panda, N., Goh, K.-S., & Chang, E. Y. (2006). Active learning in very large databases. *Multimedia Tools and Applications, 31,* 249–267. https://doi.org/10.1007/s11042-006-0043-1.
27. Demir, B., Persello, C., & Bruzzone, L. (2011). Batch-mode active-learning methods for the interactive classification of remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing, 49,* 1014–1031. https://doi.org/10.1109/tgrs.2010.207292 9.
28. Shannon, C. E. (2001). A mathematical theory of communication. *ACM SIGMOBILE Mobile Computing and Communications Review, 5,* 3–55. https://doi.org/10.1002/j.1538-7305.1948.tb013 38.x.
29. Fu, W., Hao, S., & Wang, M. (2016). Active learning on anchorgraph with an improved transductive experimental design. *Neurocomputing, 171,* 452–462. https://doi.org/10.1016/j.neucom.2015.06.046.
30. Tong, S., & Koller, D. (2001). Support vector machine active learning with applications to text classification. *Journal of Machine Learning Research, 2,* 45–66. https://doi.org/10.1162/153244302760185243.
31. Muslea, I., Minton, S., & Knoblock, C. A. (2006). Active learning with multiple views. *Journal of Artificial Intelligence Research, 27,* 203–233. https://doi.org/10.1613/jair.2005.
32. Freund, Y., Seung, H. S., Shamir, E., & Tishby, N. (1997). Selective sampling using the query by committee algorithm. *Machine Learning, 28,* 133–168. https://doi.org/10.1023/a:1007330508534.
33. Lewis, D. D., & Catlett, J. (1994). Heterogeneous uncertainty sampling for supervised learning. In *Machine learning proceedings 1994* (pp. 148–156). Elsevier. https://doi.org/10.1016/b978-1-55860-335-6.50026-x.
34. Settles, B., Craven, M., & Ray, S. (2008). Multiple-instance active learning. In *Advances in neural information processing systems, 2008* (pp. 1289–1296).
35. Olsson, F. (2009). *A literature survey of active machine learning in the context of natural language processing*. Swedish Institute of Computer Science.
36. Hoi, S. C., Jin, R., Zhu, J., & Lyu, M. R. (2006). Batch mode active learning and its application to medical image classification. In *Proceedings of the 23rd international conference on machine learning, 2006* (pp. 417–424). ACM. https://doi.org/10.1145/1143844.1143897.
37. Rubens, N., Elahi, M., Sugiyama, M., & Kaplan, D. (2015). Active learning in recommender systems. In *Recommender systems handbook* (pp. 809–846). Springer. https://doi.org/10.1007/978-0-387-85820-3_23.
38. Zhang, C., & Chen, T. (2002). An active learning framework for content-based information retrieval. *IEEE Transactions on Multimedia, 4,* 260–268. https://doi.org/10.1109/tmm.2002.1017738.
39. Heath, M., Bowyer, K., Kopans, D., Kegelmeyer, P., Moore, R., Chang, K., & Munishkumaran, S. (1998). Current status of the digital database for screening mammography. In *Digital mammography* (pp. 457–460). Springer. https://doi.org/10.1007/978-94-011-5318-8_75.
40. USF digital mammography home page (2007). http://marathon.csee.usf.edu/Mammography/Database.html.
41. Rose, C., Turi, D., Williams, A., Wolstencroft, K., & Taylor, C. (2006). Web services for the DDSM and digital mammography research. In *International workshop on digital mammography, 2006* (pp. 376–383). Springer. https://doi.org/10.1007/11783237_51.
42. Karssemeijer, N., & te Brake, G. M. (1996). Detection of stellate distortions in mammograms. *IEEE Transactions on Medical Imaging, 15,* 611–619. https://doi.org/10.1109/42.538938.
43. Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. In *IEEE Computer Society conference on computer vision and pattern recognition, 2005. CVPR 2005* (pp. 886–893). IEEE. https://doi.org/10.1109/cvpr.2005.177.
44. Wold, S., Esbensen, K., & Geladi, P. (1987). Principal component analysis. *Chemometrics and Intelligent Laboratory Systems, 2,* 37–52. https://doi.org/10.1016/0169-7439(87)80084-9.
45. Gumus, E., Kilic, N., Sertbas, A., & Ucan, O. N. (2010). Evaluation of face recognition techniques using PCA, wavelets and SVM. *Expert Systems with Applications, 37,* 6404–6408. https://doi.org/10.1016/j.eswa.2010.02.079.
46. Liu, G., Gao, X., You, D., & Zhang, N. (2016). Prediction of high power laser welding status based on PCA and SVM classification of multiple sensors. *Journal of Intelligent Manufacturing*. https://doi.org/10.1007/s10845-016-1286-y.
47. Moura, D. C., & López, M. A. G. (2013). An evaluation of image descriptors combined with clinical data for breast cancer diagnosis. *International Journal of Computer Assisted Radiology and Surgery, 8,* 561–574. https://doi.org/10.1007/s11548-013-0838-2.
48. Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20,* 273–297. https://doi.org/10.1007/bf00994018.
49. Chang, C.-C., & Lin, C.-J. (2011). LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology, 2,* 27. https://doi.org/10.1145/1961189.1961199.
50. Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters, 27,* 861–874. https://doi.org/10.1016/j.patrec.2005.10.010.
51. Huang, H., Zhang, C., Hu, Q., & Zhu, P. (2016). Multi-view representative and informative induced active learning. In *Pacific Rim international conference on artificial intelligence, 2016* (pp. 139–151). Springer. https://doi.org/10.1007/978-3-319-42911-3_12.