Neural Comput & Applic
DOI 10.1007/s00521-014-1764-0



# Breast tumor detection in double views mammography based on extreme learning machine

**Zhiqiong Wang · Qixun Qu · Ge Yu · Yan Kang**

Received: 20 September 2013 / Accepted: 16 October 2014


**Abstract** Mammography is one of the most important methods for breast tumor detection, while existing computer-aided diagnosis (CAD) technology based on single-view mammograms ignores the contrastive feature between medio-lateral oblique (MLO) and cranio-caudal (CC) views, and CAD technology based on double-view overlooks features of single views. But in clinical environment, radiologists not only read both CC view images and MLO view images individually, but also contrast these two types of views to diagnose each case. Therefore, to simulate diagnosis process of radiologists, in this paper, a fused feature model which blends features of single views with contrastive features of double views is proposed. The fused feature model is optimized by means of feature selection methods. Then, a CAD detection method based on extreme learning machine, a classifier with wonderful universal approximation capability, is proposed to improve the effectiveness of breast tumor detection by applying the optimum fused feature. The effectiveness of proposed method is verified by 222 pairs of mammograms from 222 women in Northeast China through the complete experiment.

**Keywords** Extreme learning machine (ELM) · Computer-aided diagnosis (CAD) · Mammograms · Image processing · Feature selection

Z. Wang · Q. Qu · Y. Kang (✉)
Sino-Dutch Biomedical and Information Engineering School, Northeastern University, Shenyang, China
e-mail: kangyan@bmie.neu.edu.cn

Z. Wang
e-mail: wangzq@bmie.neu.edu.cn

Q. Qu
e-mail: quqixun@gmail.com

Z. Wang
Key Laboratory of Medical Image Computing, Ministry of Education, Shenyang, China

Z. Wang · G. Yu
College of Information Science and Engineering, Northeastern University, Shenyang, China
e-mail: yuge@ise.neu.edu.cn

## 1 Introduction

Breast cancer is one of the common malignant tumors which affects the fitness of women. Scientific evidence suggests that the early diagnosis and prediction can reduce the morbidity and mortality of the breast cancer by a big degree [1, 2]. At present, computer-aided diagnosis (CAD) of breast tumor detection based on mammograms can provide radiologists with a valuable opinion to improve the accuracy and efficiency of early-stage breast cancer detection and diagnosis [3–5]. Existing detection methods only pay attention to diagnose cranio-caudal (CC) view and medio-lateral oblique (MLO) view of mammograms individually, the others only concern about the contrastive characters between CC view and MLO view. Though the existing technologies acquire some achievements of breast tumor detection, they are needed to improve the accuracy and efficiency.

In clinical environment, radiologists read both CC view images and MLO view images to diagnose each case. They locate breast tumors in different single views of a breast individually first by observing the morphology and gray scale of breast tissue. Then, they compare relevant areas in double views of a breast to verify the diagnosis result. Thus, to describe a breast more like what radiologists do, a fused feature model of a breast should consist two distinct

Published online: 16 November 2014

parts, one is the single-view feature vector and another is the double-view contrastive feature vector. The fused feature model takes advantage of single-view features to find out all suspected tumors of a breast and determines real tumors by applying contrastive features of double views. The fused feature model enhances the accuracy of breast tumor detection.

In machine learning field, extreme learning machine (ELM) has caught much attention in recent years. ELM has been developed for single hidden layer feed-forward neural networks learning algorithm. ELM has both universal approximation and classification capabilities, because hidden neurons can be randomly generated and independent from applications as well. ELM provides better generalization performance than other conventional learning algorithms at an extremely fast learning speed, ease of implementation and minimal human intervention. Its wonderful performance on classification issues are widely recognized and applied in a lot of applications [6–17]. Thus, breast tumor detection applies ELM for higher efficiency.

Therefore, a computer-aided tumor detection based on ELM with the fused feature model of mammograms is proposed. First, the feature vector of single-view images and contrastive feature vector of double-view images are created. Next, integrating all those images feature vectors creates the fused feature vector combining the single- and double-view feature vectors. Then, three feature selection methods, impact value selection, sequential forward selection and genetic selection, are applied, respectively, to have a feature selection for the fused vectors, turning out the optimal feature vector. Finally, with the use of ELM make the detection for double-view images.

The contributions of this paper are as follows:

- A fused feature model, a clinical diagnosis simulation, of double-view mammogram is proposed, which blends features of single views with contrastive features of double views.
- A breast tumor CAD method based on ELM, a classier with wonderful universal approximation capability, is proposed to improve the effectiveness of breast tumor detection.
- The effectiveness of our method is verified through 222 pairs of mammograms from 222 women in Northeast China.

The remainder of this paper is organized as follows. Section 2 briefly introduces the related work. Section 3 introduces the details of a computer-aided tumor detection based on ELM with the combination of single- and double-view feature vectors of mammograms. In Sect. 4, we analyze the performance evaluation of breast tumor detection based on ELM. Finally, the results and conclusion are summarized in Sect. 5.

## 2 Background

### 2.1 Breast tumor detection

Mammogram is considered as the most reliable and the most effective way to diagnose the breast cancer in early stage. In the clinical environment, two kinds of single views of a breast, CC view and MLO view, are used for breast tumor detection. To acquire CC view images, patients are asked to face the imaging table, stand close to the camera and let the breast is placed on the cassette, and then X-ray is vertically projected downward. MLO view images are obtained by an additional check on the basis of axial position of breast. Existing detection methods only pay attention to diagnose CC and MLO views of mammograms individually, the others only concern about the contrast between CC view and MLO view.

#### *2.1.1 Single-view breast tumor detection*

Kegelmeyer [18] evaluates the performance of stellate lesion detection algorithm on the new mammogram data set while presenting a revision of the spatial integration step which generates the final report of a lesion's existence, one that facilitates the extraction of ROC performance statistics. Chan et al. [19] study the effectiveness of using texture features derived from spatial gray-level dependence (SOLD) matrices for classification of masses and nod breast tissue on mammogram. Karssemeijer et al. [20] present a method, based on statistical analysis of a map of pixel orientations, to detect stellate patterns, usually found out in malignant densities. Sahin et al. [21] introduce a new rubber band straightening transform (RBST) for characterization of mammographic masses as malignant or benign. Mudigonda et al. [22] attempt computer-aided classification of benign and malignant masses on mammograms by computing gradient-based and texture-based features, developed to estimate the sharpness of mass boundaries in the ribbons of pixels extracted from their margins. Sahin et al. [23] develop new computer vision techniques for characterization of breast masses on mammograms to improve our characterization method by making use of morphological features. Shi et al. [24] develop an automated method for mammographic mass segmentation and explore new image based features in combination with patient information in order to improve the performance of mass characterization.

Breast tumor detection based on single view concerns about the morphology and gray scale of breast tissue.

Existing solutions proposed many effective image processing methods and many significant descriptors for breast tumor. But these approaches are susceptible to false-positive tumors. In one view of a breast, there is a tumor, but in another view of the same breast, maybe there is no tumor. The contrast of two views of the same breast can avoid this problem.

#### *2.1.2 Double-view breast tumor detection*

To avoid breast tumor detection affected by false-positive tumors in single-view method, double-view methods have been proposed by contrasting CC view and MLO view of a breast. Engeland et al. [25] present a method to link potentially suspicious mass regions detected by a CAD scheme in MLO and CC mammographic views of the breast. Zheng et al. [26] develop and test a new multiview-based CAD scheme that aims to maintain the same case-based sensitivity level as a single-image-based scheme while substantially increasing the number of masses being detected on both ipsilateral views. Karssemeijer et al. [27] present a method which is based on a previously developed approach to linking potentially suspicious regions in MLO and CC views to improve CAD results for masses in mammograms by fusing information obtained from two views of the same breast. Sarnulsk et al. [28] present a method to match corresponding regions in MLO and CC mammographic views of the breast. Pu et al. [29] develop a method to automatically register breast areas and detect matching strips of interest used to identify the matched mass regions depicted on CC and MLO views, using an ellipse-based model to fit the breast boundary contour (skin line) and setting a local Cartesian coordinate system for each view.

Breast tumor detection based on double views pays attention to the contrast characteristic of CC and MLO views of a breast. Existing methods provide good approaches to matching suspicious regions in two views images and offer many valuable features to describe the relations between matched regions. However, double-view detection ignores the single-view features, which results in omissions of real tumors. Therefore, the tumor detection based on the integration of single-view method and double-view method becomes the emphasis and orientation of feature research. The fused feature model is proposed which combines single-view feature and contrastive features of double views to improve the accuracy of breast tumor detection.

### 2.2 Extreme learning machine

ELM has been developed for single hidden layer feedforward neural networks learning algorithm. The theories of ELM [6–9] show that it has both universal approximation and classification capabilities, because hidden neurons can be randomly generated and independent from applications as well. ELM provides better generalization performance than other conventional learning algorithms at an extremely fast learning speed, ease of implementation and minimal human intervention. Compared with SVM or other classifiers based on SVM such as LS-SVM and PSVM, ELM is a unified learning platform which can deal with a wide range of feature mappings and can provide for regression and multiclass classification; ELM can be optimized under a milder optimization constraints than SVMs; ELM can estimate any target continuous function and classify any separate categories [9]. ELM has been appliedto plenty of applications and studies, which is followed by a number of variants of basic ELM, for instance, fully complex ELM, online sequential ELM, incremental ELM, pruning ELM [10–17]. In the paper, the basic ELM is applied.

## 3 Breast tumor detection based on ELM

### 3.1 General framework

The process to detect mammary tumor of the isolated mammography has four steps: The first one is image preprocessing of mammogram, including denoising, enhancement, edge segmentation feature extraction and registration to acquire the position of tumor and its boundary. The next part is the building of feature model. It is a five-step procedure involving feature extraction of single view, feature extraction of double views, fusions of two views and feature selection. Feature extraction of single view has covered morphological and texture feature. Contrastive feature extraction of double views has covered the invariance and similarity feature of groups of MLO view and CC view. Fusions of two views mean the feature combination of features in the single view and the double view. The acquisition of feature vector is implemented through the optimization of feature sets on the impact value selection, sequential forward selection and genetic algorithm selection. By now, there are four feature models established. At last, classification based on ELM is executed with the use of cross validation between four feature models.

### 3.2 Mammogram processing

The main procedure of image preprocessing is the denoising with median filter and enhancement of contrast. The result of preprocessing is shown in Fig. 1. Next step is the possession for positions of chest wall, axle wire and nipple in MLO view and possession for positions of axle wire and

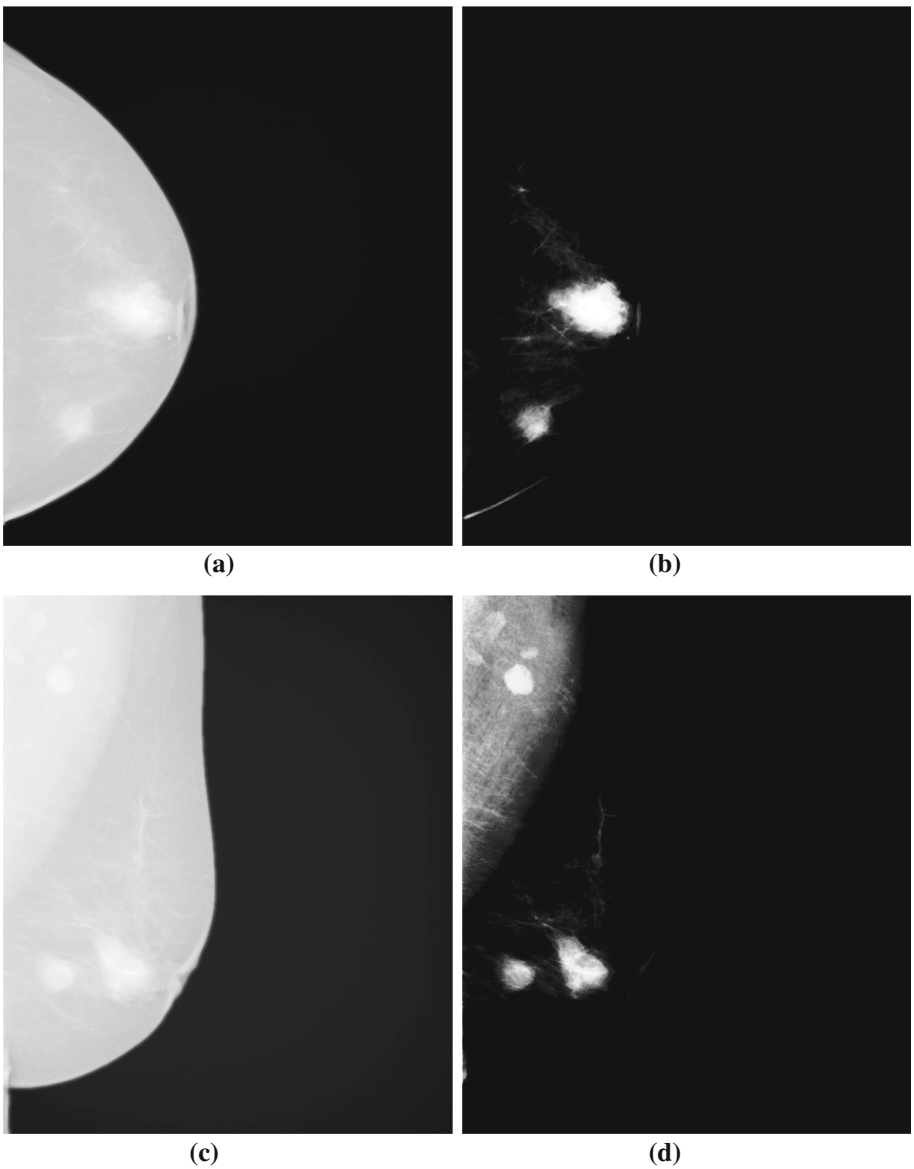


**Fig. 1** Contrast images of before and after preprocessing. **a** Original CC view image, **b** after preprocessing of **a**, **c** original MLO view image, **d** after preprocessing of **c**

nipple in CC view to find the ROI with suspicious tumor in two views, as shown in Fig. 2. Then, the application of wavelet modulus maxima algorithm and morphological operation is utilized to segment possible tumor within ROI for tumors boundaries, which can be seen in Fig. 3. The following procedure should be spline region registration to register tumors. The tumors boundaries and result from registration of tumor should be applied with the feature extraction.

## 3.3 Single-view feature vector building

### *3.3.1 Single-view feature vector*

Features of an image reveal the basic attributions and characteristics. Therefore, the parameters of features should be recognizable, reliable and independent. A model of single-view feature vector has been established [33]. There are five geometry features of a tumor, and they are roundness, entropy of standardized radius, variance of standardized radius, ratio of area and roughness. There are five features of texture based on GLCN, and they are reverse gap, entropy, energy, correlation coefficient and contrast, as shown in Table 1.

### *3.3.2 Geometry features extraction*

The breast tumors are variance in shape, and its location is irregular. Because of the variance, the roughness, size, shape, edge and density are geometrical features that are important basis of tumor detection. The tumor can be interpreted as a light spot on medical images. The

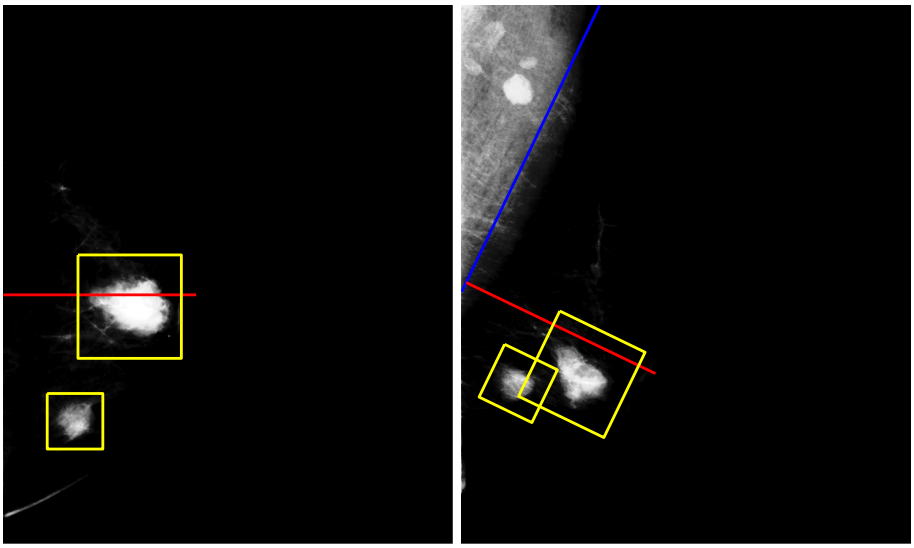


**Fig. 2** Get ROIs from two views of mammograms. **a** CC view, **b** MLO view

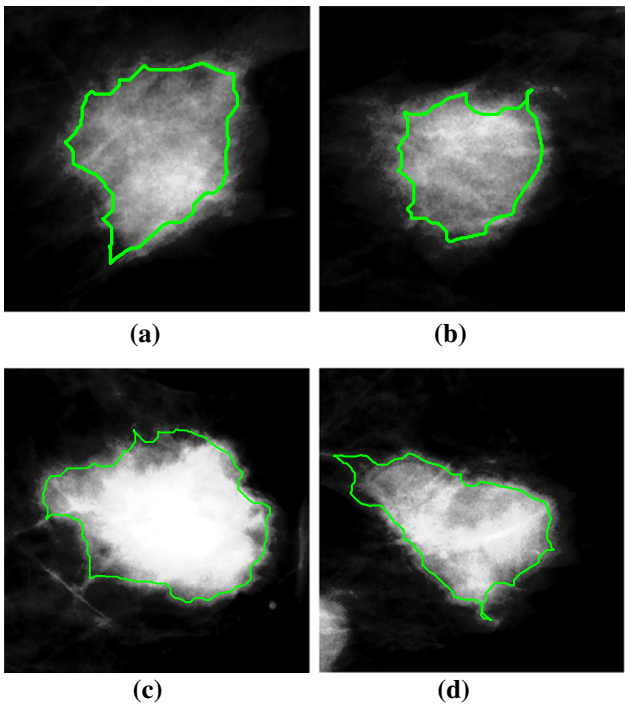


**Fig. 3** Edges of tumor in two views of mammograms. **a** First tumor in CC view, **b** first tumor in MLO view, **c** second tumor in CC view, **d** second tumor in MLO view

grayscale value is higher than neighbor regions, and the grayscale value of edge is a transition. The values of gray scale are decreased with neighbor tissues.

The common shape of a breast tumor includes circle, ellipse, and most of them is benign, but they are irregular shaped, blurred and star sized that are malignant. Based on the segment of tumor, we extract and analyze its geometrical features, including roundness, entropy of standardized radius, variance of standardized radius, ratio of area and roughness.

**Table 1** Single-view features

| Feature type | Name | Feature expression |
|---|---|---|
| Geometry features | Roundness | $g_1 = \frac{P^2}{A}$ |
| | Entropy of standardized radius | $g_2 = -\sum_{k=1}^{100} p_k(\log(p_k))$ |
| | Variance of standardized radius | $g_3 = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(d(i) - d_{\text{avg}})^2}$ |
| | Ratio of area | $g_4 = \frac{1}{d_{\text{avg}}N}\sum_{i=1}^{N}(d(i) - d_{\text{avg}})$ |
| | Roughness | $g_5 = \frac{1}{N}\sum_{i=1}^{N}\lvert d(i) - d(i+1)\rvert$ |
| Texture features | Inverse difference moment | $t_1 = \sum \frac{P(i,j)}{1+(i-j)^2}$ |
| | Entropy | $t_2 = \sum P(i,j) \times [-\ln P(i,j)]$ |
| | Energy | $t_3 = \sum P^2(i,j)$ |
| | Correlated coefficient | $t_4 = \sum \frac{P(i,j)\times(i-\mu_x)\times(j-\mu_y)}{\delta_x\delta_y}$ |
| | Contrast | $t_5 = (i-j)^2 \times P(i,j)$ |

*Roundness* This is one of the most important geometrical features. The smoother the edge is, the smaller the roundness is, and the greater the odd of benign tumor is. In contrast, the bigger the roundness is, the greater the odd of the malignant tumor is. The Roundness is defined in Eq. 1.

$$g_1 = \frac{P^2}{A} \tag{1}$$

where $A$ is the area of tumor and $P$ is the girth of edge.

*Entropy of standardized radius* It means the difference between standardized radii. The definition is in Eq. 2.

$$g_2 = -\sum_{k=1}^{100} p_k(\log(p_k)) \tag{2}$$

where $p_k$ is the probability of standardized histogram. The standardized histogram is percentage of division into 100 pieces between the biggest radius and the smallest radius to calculate the number of radii within them. It means the $k$th percentage of radius.

*Variance of standardized radius* It describes the range of variance of standardized radius. It is defined in Eq. 3.

$$g_3 = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\left(d(i) - d_{\text{avg}}\right)^2} \tag{3}$$

where $N$ is the number of edge points, $d(i)$ is the $i$th standardized radius of edge points.

*Ratio of area* It describes the circularness of a tumor. If the value is smaller, the shape of the image is similar to a circle. The ratio of area is defined in Eq. 4.

$$g_4 = \frac{1}{d_{\text{avg}}N}\sum_{i=1}^{N}\left(d(i) - d_{\text{avg}}\right) \tag{4}$$

where $d(i) - d_{\text{avg}} = 0 \forall d(i) \le d_{\text{avg}}$ is the average standardized radius of edge points.

*Roughness* The more irregular the shape is, the bigger the roughness. It is defined in Eq. 5.

$$g_5 = \frac{1}{N}\sum_{i=1}^{N}|d(i) - d(i+1)| \tag{5}$$

### 3.3.3 Texture features extraction

The appearance of tumor and calcification twists the image of breast tissues. As a result, texture feature becomes significant in the detection of tumor. It is also a reliable basis of classification. Currently, there are many algorithms for texture feature extraction. The popular one is the texture extraction algorithm based on gray-level co-occurrence matrix (GLCM). GLCM can reflect the information of pixels during relative locations. It can also reflect the properties of distributions of different gray scale in relatively space. It is one of the important features in feature extraction.

The element of GLCM $(i,j)$ is the coherent distribution for probability of a gray scale $(i,j)$ in constant distance $d$. When scanning the image data from left to right, up to down, the GLCM is constructed by the reflection of co-occurrence time probability of related pixels in space. The GLCM shows the joint probability distribution of two grayscale pixels with distance $(\Delta x, \Delta y)$.

If the grayscale level of a image data is $N$, the GLCM will be a $N \times N$ matrix. It can be shown in the form of $M(\Delta x, \Delta y)(h,k)$, where value of $Mhk$ at $(h,k)$ is the counting number of pixel $h$ and pixel $k$ in the distance of $(\Delta x, \Delta y)$. $f(x,y)$ is a digital image data with the size of $M \times N$ and grayscale level $N_{\text{g}}$. Then, the definition of space-conditional GLCM is in Eq. 6.

$$\begin{aligned} P(i,j) = \#\{(x_1,y_1),(x_2,y_2) \in M \times N | f(x_1,y_1) \\ = i, f(x_2,y_2) = j\} \end{aligned} \tag{6}$$

where $\#(x)$ is the number of elements in the set $x$. It is obvious that $P$ is a $N_{\text{g}} \times N_{\text{g}}$ matrix. The distance between $(x_1,y_1)$ and $(x_2,y_2)$ is assumed in $d$, and the angle between coordinate axis is $\theta$. Therefore, GLCM $P(i,j,d,\theta)$ for any distances and angles is obtained.

The number of parameters of texture feature extraction of segmented tumor based on GLCM is five, including inverse difference moment, entropy, energy, correlated coefficient and contrast.

*Inverse difference moment* It reveals the homogeneity of texture in an image to show the extent of change in texture. The bigger the value is, the tinier the change is in texture. It is defined by Eq. 7.

$$t_1 = \sum \frac{P(i,j)}{1 + (i-j)^2} \tag{7}$$

where $p(i,j)$ is the element of row $i$ and column $j$ of GLCM.

*Entropy* It deals with the amount of information in an image. Texture is one of the information in image. Besides, texture has random property. If co-occurrence matrix has the biggest random property, values of elements in GLCM are equal approximately and distributed varyingly, which yields a bigger value of entropy. It reveals the extent of uniformity and complexity. It is defined in Eq. 8.

$$t_2 = \sum P(i,j) \times [-\ln P(i,j)] \tag{8}$$

*Energy* It is the sum of square of every values of gray scale in GLCM. Consequently, it reveals the fineness of texture and uniformity of grayscale distribution. If every value in GLCM is equivalent, values of energy are the lowest. In contrast, if every value in GLCM are uneven, values of energy are bigger. Energy is the value that shows the mode of uniform change in texture. The definition of it is in Eq. 9.

$$t_3 = \sum P^2(i,j) \tag{9}$$

*Correlated coefficient* This parameter is the measurement of similarities on the direction of column and row. If elements in GLSM are uniformly equivalent, correlated coefficient will become bigger. Otherwise, it is a smaller

value. If the horizontal texture exists, the value of this direction is bigger than other direction. The definition of it is in Eq. 10.

$$t_4 = \sum \frac{P(i,j) \times (i - \mu_x) \times (j - \mu_y)}{\delta_x \delta_y} \tag{10}$$

where $\mu_x, \mu_y, \delta_x, \delta_y$ are the average and variance of $P_x, P_y$, and the definition of $P_x, P_y$ is in Eq. 11.

$$P_x = \sum_{j=1}^{N_g} P(i,j), P_y = \sum_{i=1}^{N_g} P(i,j) \tag{11}$$

*Contrast* It reflects the extent of images textures and the degree of difficulty. If the texture is deep, the contrast will become bigger to yield good visual effects. Otherwise, visual effects will be dissatisfactory. The significant difference between grayscale means notable elements with big contrast. The value of contrast depends on the values of elements that are away from the diagonal of matrix. The definition of contrast is in Eq. 12.

$$t_5 = (i - j)^2 \times P(i,j) \tag{12}$$

### 3.4 Contrastive feature vector building

#### *3.4.1 Contrastive feature vector*

Contrastive feature vector of double views is defined in Eq. 13.

$$df = [df_1, df_2] \tag{13}$$

where $df_1$ and $df_2$ are multidimensional vectors, which, respectively, represent the invariance feature of related views and the feature of similarity measurement, and their mathematical equations are as follows:

$$df_1 = [I_1, I_2, I_3] \tag{14}$$

$$df_2 = [S_1, S_2, S_3, S_4, S_5] \tag{15}$$

of integrated density of the related tumors and local binary pattern texture analysis feature. And the five features of similarity measurement are $S_1$, $S_2$, $S_3$, $S_4$ and $S_5$, which, respectively, represent joint entropy, conditional entropy, average Kullback–Leibler (KL) divergence, maximum KL divergence and arithmetic–geometric mean divergence.

#### *3.4.2 Invariance features extraction*

When the segmentation of tumor is completed, the feature of the tumors CC view and MLO view images would be extracted. However, the same tumor in different views expresses wide difference because of the diversity, irregular and asymmetry of the breast tumor. Furthermore, when picturing the mammography, the breast would be press in some extent, which can affect the detection accuracy. So, it is necessary to extract the feature that would not be affected by factors above, which is called the invariance feature of related views.

In this paper, the absolute value of distance between the projection point and nipple in the registration strip of MLO and CC views is extracted as one of the invariance features of related views. Because it has a high correlation between the suspicious area and the nipple, the integrated density [30] of the related tumors is also be extracted. Because it has nothing to do with projection perspective and deformation of the tissue, its invariance in different views is very high. A texture analysis feature named local binary pattern (LBP) [31] is extracted, which has the invariance in rotation. Three related features are provided with the possibility distribution histogram of LBP modality in the tumor region of CC and MLO views. And these three feature are the other three the invariance feature of related views based on the invariance in rotation of LBP. The computational formulas of the three features are as follow:

$$I_1 = 1 - \frac{1}{2} \sum_{i=0}^{9} \left| \frac{H_{\mathrm{CC}}(\mathrm{LBP_i})}{\mathrm{Num_{CC}(LBP)}} - \frac{H_{\mathrm{MLO}}(\mathrm{LBP_i})}{\mathrm{Num_{MLO}(LBP)}} \right| \tag{16}$$

$$I_2 = \frac{\sum_{i=0}^{9} \left( H_{\mathrm{CC}}(\mathrm{LBP_i}) - \overline{\mathrm{H_{CC}(LBP)}} \right) \left( H_{\mathrm{MLO}}(\mathrm{LBP_i}) - \overline{\mathrm{H_{MLO}(LBP)}} \right)}{\sqrt{\sum_{i=0}^{9} \left( H_{\mathrm{CC}}(\mathrm{LBP}_i) - \overline{H_{\mathrm{CC}}(\mathrm{LBP})} \right)^2} \sqrt{\sum_{i=0}^{9} \left( H_{\mathrm{MLO}}(\mathrm{LBP}_i) - \overline{H_{\mathrm{MLO}}(\mathrm{LBP})} \right)^2}} \tag{17}$$

where $I_1$, $I_2$ and $I_3$ represent the invariance features of related views, which, respectively, are the absolute value of distance between the projection point and nipple in the registration strip of MLO and CC views, the absolute value

$$I_3 = \sum_{i=0}^{9} \frac{(H_{\mathrm{CC}}(\mathrm{LBP_i}) - \mathrm{H_{MLO}(LBP_i)})^2}{H_{\mathrm{CC}}(\mathrm{LBP}_i) + H_{\mathrm{MLO}}(\mathrm{LBP}_i)} \tag{18}$$

where $I_1$, $I_2$, $I_3$, respectively, represent the normal correlation of LBP modal distribution histogram, Pearson correlation and the correlation analysis of $(\chi^2)$ in the registration area of MLO and CC views. Where $H_{\mathrm{CC}}(\mathrm{LBP}_i)$ and $H_{\mathrm{MLO}}(\mathrm{LBP}_i)$, respectively, are the occurrence number of the LBP modality in the registration region of the CC view and MLO view images. $\mathrm{Num}_{\mathrm{CC}}(\mathrm{LBP})$ and $\mathrm{Num}_{\mathrm{MLO}}(\mathrm{LBP})$, respectively, are total occurrence number of LBP modality in the registration region of the CC view and MLO view images. $\overline{H_{\mathrm{CC}}(\mathrm{LBP})}$ and $\overline{H_{\mathrm{MLO}}(\mathrm{LBP})}$, respectively, are the average value of all the kinds of LBP modality occur in the images that view in CC and MLO views.

#### 3.4.3 Similarity features extraction

In the article, we extract five features of similarity measurement based on the theory of information as the correlation feature of the matching area in CC and MLO images. Relatively are joint entropy, conditional entropy, average KL divergence, maximum KL divergence and arithmetic–geometric mean divergence [32].

*Joint entropy* Joint entropy represents the entropy of the joint histogram of image $X$, $Y$. If images $X$, $Y$ are completely irrelevant, the sum of their joint entropy is equal to their individual entropies. On the conversely aspect, the more similar $X$ and $Y$ are, the lower their joint entropy is compared with the sum of the individual entropies, which is as shown in Eq. 19.

$$S_1 = H(X,Y) = -\sum_x p(x,y)\log_2\left[P_{xy}(x,y)\right] \tag{19}$$

$$H(X) = -\sum_x p(x)\log_2[P_x(x)] \tag{20}$$

where $H(x)$ is Shanon entropy and $p(x)$ is the probability that an image pixel will have the intensity value $x$.

*Conditional entropy* The conditional entropy $H(X \mid Y)$ of two images $X$ and $Y$ measures how much entropy (or uncertainty) is remaining regarding image $X$ when the truth regarding image $Y$ has learned. Its form of mathematics is shown in Eq. 22.

$$S_2 = H(X \mid Y) = H(X,Y) - H(Y) \tag{21}$$

With conditional entropy, the conditional entropy (or uncertainty) of the query image given the known image should be low if two images are relevant.

*Average and maximum Kullback–Leibler divergence* KL divergence is a distance measure between two probability distributions $p(x)$ and $q(x)$ which, respectively, are the possibility distributions of the stored image $p(x)$ and the query image $q(x)$. The computational formula of KL divergence is in Eq. 22:

$$D(q \parallel p) = \sum_x q(x)\log_2\left[\frac{q(x)}{p(x)}\right] \tag{22}$$

Normally, the higher KL divergence when the two images are more dissimilar can be achieved. And the two transformations, average KL divergence and maximum KL divergence, respectively, are as follows:

$$S_3 = \frac{D(q \parallel p) + D(p \parallel q)}{2} \tag{23}$$

$$S_4 = \max[D(q \parallel p), D(p \parallel q)] \tag{24}$$

*Arithmetic–geometric mean divergence* This measure is essentially the KL divergence between the arithmetic and geometric means of the two image distributions $p(x)$ and $q(x)$. The computational formula is in Eq. 25:

$$S_5 = \sum_x \frac{p(x)+q(x)}{2}\log_2\frac{p(x)+q(x)}{2\sqrt{p(x)q(x)}} \tag{25}$$

### 3.5 Features vector fusion

Doctors usually examine the different views of mammography in clinical diagnosis. And the single mammography and double mammography with MLO and CC views can provide different information, which are valuable for the diagnosis. Therefore, detection with single mammography and double mammography through features fusion can improve the accuracy of this CAD system.

The fused features are extracted not only from the single-view mammograms, but also from the double-view mammograms. A model of fused feature vector is established mathematically in Eq. 26.

$$d = [df, sf_{\mathrm{MLO}}, sf_{\mathrm{CC}}] \tag{26}$$

where $sf_{\mathrm{MLO}}$ is the single-view feature vector of MLO view of mammograms and $sf_{\mathrm{CC}}$ is the single-view feature vector of CC view of mammograms. And $df$ is the fused feature vector of MLO and CC views of mammograms.

### 3.6 Features selection

There are too many features that can describe the tumor in the image, yet they do not all independently represent the tumor. It is because that some features are dependent on other features, that is say that some features can determent some other features value. So, when we extract these features and input them into ELM, and it can make ELM over learned easily. Then, it affects the result of ELM learning progress and decreases the accuracy of tumor detection. So that, it is significant to optimize the feature vector that we extract. There are many algorithms that can select

features [34] raise three feature selections that, respectively, are impact value selection, sequential forward selection and genetic algorithm selection.

*Impact value selection* The main idea of impact value selection (IVS) is that eliminate the redundant feature, which is not the pivotal feature and affect the classification less; otherwise preserve the key feature that determine the result. The pseudo-code is in Algorithm 1.

**Algorithm 1:** Impact Value Selection

```
1  // input OFM: Original feature vector model
2  // output NFM: New feature vector model
3  // newAcc: Result of classification
4  oldAcc=ELM(OFM);
5  For i = 1 to N Do
6      OFMcopy = OFM;
7      OFMcopy(i) = 0;
8      acc = ELM(OFMcopy);
9      If |acc − oldAcc| > D
10         AppendTo(NFM, OFM(i));
11 newAcc = ELM(NFM);
12 Return ⟨NFM, newAcc⟩;
```

For given the feature vector model that contains $N$ features, the precision of classification based on this model is calculated (Line 4). The diversity factor of some feature is set to no difference (Line 6, 7). Calculate the precision of classification with the new feature vector model (Line 8). The absolute value of the difference between the new precision and the old one with the certain feature is defined as the influential value. If the value is larger than the threshold $D$, it will be defined as the key feature; otherwise, it is defined as the redundancy feature (Line 9, 10). Along this procedure, input the combination of all the key features into the ELM and achieve the best classification result.

*Sequential forward selection* Sequential forward selection (SFS) is a simple greedy algorithm. When process the algorithm, calculating the classification capacity of each feature, and then assemble the features that has the most strong classification capacity to a best feature vector. Then recalculate the vector until the best classification feature vector has been achieved. The pseudo-code is as Algorithm 2.

**Algorithm 2:** Sequential Forward Selection

```
1  // input OFM: Original feature vector model
2  // output NFM: New feature vector model
3  // newAcc: Result of classification
4  For i = 1 to N Do
5      sAcc = ELM(OFM(i));
6      AppendTo(allsAcc, (sAcc, i));
7  sortAcc=sort(allsAcc, sAcc);
8  For i = 1 to N Do
9      AppendTo(FM, OFM(sortAcc(i)));
10     AppendTo(allAcc, ELM(FM));
11 newAcc = FindMax(FM);
12 Return ⟨NFM, newAcc⟩;
```

For given the feature vector model that contains $N$ features, input every single into the ELM and calculate their classification ability (Line 4–6). Then, ordering all features based on their classification ability (Line 7). Start with the empty feature vector, add the feature into the vector one by one along their order and calculate the latest precision of the new vector. The previous procedure will be repeated until all the features are added into the vector and form the vector with certain number of feature (Line 8–10). When the best classification vector that contains certain number of feature is achieved, all the features in the vector are key feature (Line 11–13).

*Genetic algorithm selection* Genetic algorithm selection (GAS) is a progress of random search. It simulates the procedure of natural selection and heredity in nature. And it aims to find the most adaptability feature vectors which exactly are the key feature to classification, which are shown in Algorithm 3.

**Algorithm 3:** Genetic Algorithm Selection

```
1  // input OFM: Original feature vector model
2  // output NFM: New feature vector model
3  // newAcc: Result of classification
4  initPop = rand(OFM, n);
5  tPop = initPop;
6  For i = 1 to G Do
7      For j = 1 to n Do
8          If (fitness(tPop(j)) > S)
9              AppendTo(nPop, tPop(j));
10         tPop = nPop; Empty(nPop);
11         newPop(tPop);
12 NFM = FindMax(tPop);
13 newAcc = ELM(NFM);
14 Return ⟨NFM, newAcc⟩;
```

First of all, the initial population with $N$ individuals is generated, which represents the feature model, and the classification ability of it is calculated (Line 4). Then, the fitness of individual based on this vector is calculated, which is the reciprocal value of variance between the test result and the real result based on some combination of the feature vector. When the fitness value larger than the given threshold S, it means that the vector contains some key features and it can be the parent generation; otherwise, it cannot be the parent generation and continue inheriting (Line 5–11). In the next, the parent generates new offspring through crossover and mutation. And then continue calculating the fitness of every individual in order to find the heritable parent (Line 12). Along this procedure, a most adaptable population can be achieved, which represents the best classification ability feature vector, after evolution of $G$ populations (Line 13, 14).

### 3.7 Breast tumor detection approach

#### *3.7.1 ELM superiority*

ELM, the basic version, has both universal approximation and classification capabilities, because hidden neurons can be randomly wonderful generated and independent from applications as well. ELM provides better generalization performance than other conventional learning algorithms at an extremely fast learning speed, ease of implementation and minimal human intervention. ELM has been applying to plenty of applications and studies, which is followed by a number of variants of basic ELM.

#### *3.7.2 ELM training*

Before the ELM training which is for breast tumor detection is started, the matrix of features should be established. Algorithm 4 shows that how the function FE extracts the feature matrix. An image set which has N-pair images is used in this process. First step is reading a pair of images that are different views of the ipsilateral breast (MLO and CC view) (Line 2). When the images are loaded, they are preprocessed by the denoising and the enhancement (Line 3). Then, the segmentation is applied in two images and get tumors that can be matched in different views (Line 3,4). Next, extract five geometric features and the five texture features of every single tumor, as well as eight related features of a pair of tumors (Line 5–7). Finally, mix all features together in matrix **F**. Thus, the matrix of features is obtained (Line 8).

**Algorithm 4:** Features Extraction Function $FE$

1 **For** $i = 1$ to $N$ **Do**
2 Read images imgMLO and imgCC from a pair of images;
3 Pre-processing and segmentation on imgMLO and imgCC;
4 Get and match tumors from two views images;
5 Get geometric features from every tumor;
6 Get texture features from every tumor;
7 Get related features from a pair of tumors;
8 Mix all features together in **F**;
9 **Return** $F$;

The ELM training function FT of tumor detection is as shown in Algorithm 5. First, the number of hidden layer nodes $L$ is determined by experiment. $\mathbf{w_i}$ and $b_i$, two parameters of hidden layer nodes, are generated randomly (Line 2). Then matrix **H** is calculated according to **w** and $b$ by taking advantage of the image feature matrix (Line 3). The last step to finish the training of ELM is the obtainment of $\beta$ with the usage of **H** and doctors diagnosis result **T** (Line 4).

**Algorithm 5:** Training Function $FT$

1 **For** $i = 1$ to $L$ **Do**
2 Randomly generate hidden node parameters $(w_i; bi)$;
3 Calculate the hidden layer output matrix **H**;
4 Calculate the output weight vector $\beta = \mathbf{H}^{\dagger}\mathbf{T}$;
5 **Return**$\langle \mathbf{w}; b; \beta \rangle$;

#### *3.7.3 Classification*

Parameters **w**, $b$ and $\beta$ are obtained by ELM training, which means that the classification of ELM for breast tumor detection has been determined. The process of diagnosis is as shown in Algorithm 6. When a pair of images or a group of images is input, the first step is feature extraction as shown in Algorithm 1, and then the calculation of the **H** according to **w** and $b$ (Line 1) get diagnosis result **T** by applying the in the feature vector **F** last step (Line 2).

**Algorithm 6:** Testing Function $FC$

1 Calculate the hidden layer output matrix **H** use **F**, **w** and $b$;
2 Get classification result $\mathbf{T} = \mathbf{H} \cdot \beta$ ;
3 **Return T**;

## 4 Performance evaluation

### 4.1 Experiment settings

#### *4.1.1 Experiment data*

There are 222 pairs of mammograms in the images library, of which 112 have tumors. The patients are women from Northeast China between ages 32 and 74. Images are acquired by Senographe 2000D Full digital mammography camera during from August 17, 2005 to March 15, 2007. All the images are confirmed by doctors from Tumor Hospital of Liao Ning Province.

#### *4.1.2 Experiment design*

To improve the learning efficiency and accuracy, feature selection has been used in the experimental design for ELM. At first, three features models, which are single-view feature model, contrastive feature model and fused feature model, are applied to train and test ELM. To guarantee that a generalized performance of ELM can be received, cross validation is employed [33]. Then, get the result and all evaluation indices of the best features model that provides the highest classification performance. Next, randomly select 176 data from the best features model as the training

data TrainSet, and the others are testing data TestSet. TrainSet and TestSet have all features. Then, the methods of genetic algorithm, impact value selection and sequential forward selection are applied in the feature selection, respectively, using the same TrainSet and TestSet. Then, determine arguments for three feature selection methods. The threshold $D$ in IVS is set to 5 % after several test experiments. The evolution generation $G$ is set to 100, and the threshold $S$ is set to 80 %. Finally, the best optimized features model and its all indices are obtained.

### 4.2 Evaluation indices

Objective evaluation indices are significant to evaluate and determine a diagnosis system capability. In this paper, three kinds of evaluation profile, errors table, ROC curve and other five indices, are employed to estimate the diagnosis capability. The error table shows every group in cross validation of different feature model. The ROC cure is a graphical plot which illustrates the performance of a binary classifier system. The five indices of diagnosis capability evaluation indices are accuracy, sensibility, particularity, TPRatio, TNRatio and training time as shown in Table 2. The following parameters will be used to compute indices:

- TP: The number of accurately classified images which have tumors.
- TN: The number of accurately classified images which have no tumors.
- FP: The number of inaccurately classified images which have tumors.
- FN: The number of inaccurately classified images which have no tumors.

The higher accuracy of classification means that the learning and classification capability of ELM is better. Sensibility and TNRatio represent the ability of detecting the existence of tumors. The higher value states the more sensitive detection to existence of tumors. Particularity and TPRatio indicate the capability of detecting inexistence of tumors. The higher value makes the more sensitive detection to inexistence of tumors. The shorter the training time it takes, the higher the efficiency of ELM achieved.

**Table 2** Evaluation indices

| Indices | Calculation | Meaning of higher value |
|---|---|---|
| Accuracy | $\frac{TP+TP}{TP+TP+TP+TP}$ | Better classified capability |
| Sensibility | $\frac{TP}{TP+TP}$ | More sensitive to existence of tumors |
| Particularity | $\frac{TP}{TP+TP}$ | More sensitive to nonexistence of tumors |
| TPRatio | $\frac{TP}{TP+TP}$ | More sensitive to existence of tumors |
| TNRatio | $\frac{TP}{TP+TP}$ | More sensitive to nonexistence of tumors |

### 4.3 Experiment results

The experiment results contain two part, the first is the result of the classificatory efficiency of different feature models, which are single-view feature model, contrastive feature model and fused feature model. The consequence of this part is that the fused feature model gives the best classification outcome with ELM. The second part is the situation of optimizing fused feature model. In this part, the fused feature model is optimized by three feature selection methods such as impact value selection, sequential forward selection and genetic selection. The result shows that the fused feature model optimized with genetic selection performs best in three feature selection situations.

#### *4.3.1 Result of three features models*

This part compares the classificatory performance of three different feature models, which are single-view feature model, contrastive feature model and fused feature model. The classification results are displayed as errors table, histogram and ROC curves, respectively. The histogram presents the accuracy, sensibility, particularity, TPRatio and TNRatio of sort result of one features model.

Table 3 reveals the false numbers of all groups in cross validation with three feature models. It is observed that the fused feature model provides the fewest mistakes. Figure 4 displays all indices of every model. From both Table 3 and Fig. 4, the sensibility and TNRatio of single-view feature model is greater than its particularity and TPRatio, which indicates single-view feature model has a better ability to find out the real breast tumor. The particularity and TPRatio of contrastive feature model is better than its sensibility and TNRatio, which means contrastive feature model can better discerning false-positive tumors. The

**Table 3** Errors in every group in cross validation of different feature combination

| Group | | No. 1 | No. 2 | No. 3 | No. 4 | No. 5 | Total |
|---|---|---|---|---|---|---|---|
| Single-view features | Y | 2/21 | 5/28 | 4/20 | 5/19 | 3/24 | 19/112 |
| | N | 2/23 | 1/16 | 3/24 | 4/25 | 4/22 | 14/110 |
| | T | 4/44 | 6/44 | 7/44 | 9/44 | 7/46 | 33/222 |
| Contrastive features | Y | 7/21 | 6/28 | 2/20 | 4/19 | 4/24 | 23/112 |
| | N | 4/23 | 3/16 | 8/24 | 5/25 | 9/22 | 29/110 |
| | T | 11/44 | 9/44 | 10/44 | 9/44 | 13/46 | 52/222 |
| Fused features | Y | 3/21 | 6/28 | 2/20 | 3/19 | 3/24 | 17/112 |
| | N | 2/23 | 1/16 | 1/24 | 3/25 | 4/22 | 11/110 |
| | T | 5/44 | 7/44 | 3/44 | 6/44 | 7/46 | 28/222 |

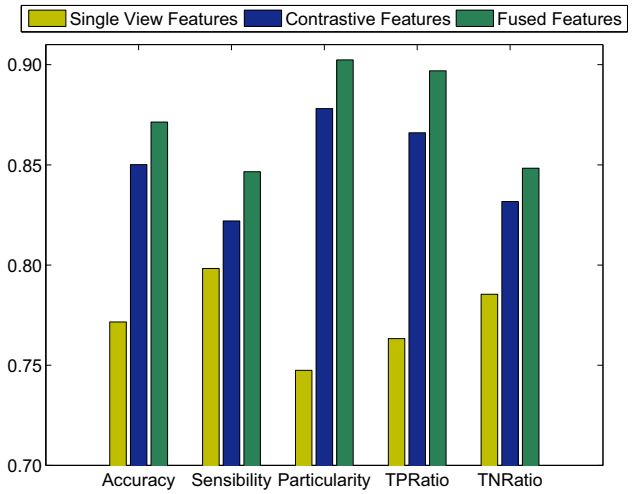


**Fig. 4** Indices of different feature selection approaches

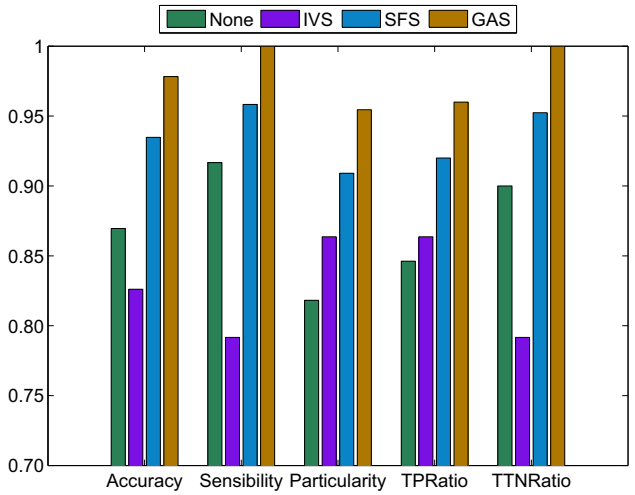


**Fig. 6** Indices of different feature combination

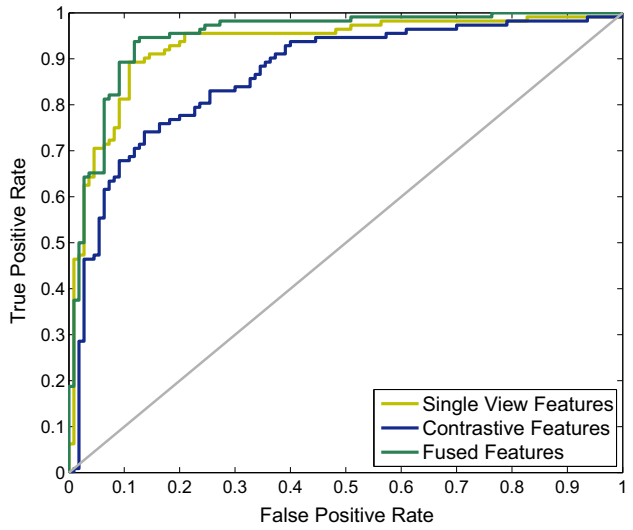


**Fig. 5** ROCs of different feature selection approaches

fused feature model gives the best result of every index, and it has pretty good recognition capability for mammograms with breast tumors or without breast tumor. Figure 5 demonstrates all ROC curves of combinations. It can be perceived that fused feature model has a larger area under the curve. Also, it is closer to the top left corner than other feature combinations. Any of the above-mentioned evidence states clear that the fused feature model of all views is much more suitable for classifying mammograms into group with tumors and group without tumors.

The fused feature model has two primary segments: features of single views and features of double views. Single-view features are extracted from detected tumors of MLO and CC views, which contain morphological features and texture features of every tumor. These two types' features are significant to estimate whether the detected tumor is true positive. Because breast tumor has a biggish size and a clear boundary, which means shape of detected tumor cannot be ignored as well as its gray scale. Double-view features are extracted by matching corresponding areas of detected tumors both in MLO view and in CC view of a same breast. If a detected tumor appears on MLO view, but there is not a tumor shown on the corresponding region of CC view, it indicates that the detected tumor is preferred as false-positive. It is intended to reduce the false-positive rate by with double-view features. So that, the fused feature model take the advantages of single-view features and double-view features to improve the accuracy of breast tumor detection.

### *4.3.2 Result of optimized features models*

To void the overfitting of ELM classification, three frequently used methods, such as impact value selection, sequential forward selection and genetic selection, are applied to select features from all views features model. The indices histogram and ROC curves present the classification performance of the fused feature model and three optimized features models.

Figure 6 shows the testing accuracy of every selection methods. It can be seen actually that genetic algorithm selection gives the highest testing accuracy than other methods. At the same time, it has the best results of indices. That means ELM gives the best effect of classifying using the feature combination that is provided by genetic algorithm selection. Figure 7 also gives evidence to support the same results.

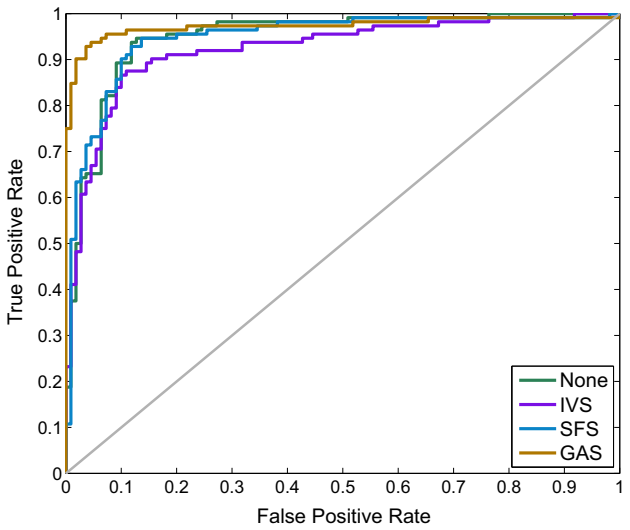


**Fig. 7** ROCs of different feature combination

## 5 Conclusion

A fused feature model which blends features of single views with contrastive features of double views is proposed to simulate radiologists' diagnosis process. The fused feature model is optimized by means of genetic selection methods. Then, a CAD detection method based on ELM is proposed to improve the effectiveness of breast tumor detection by applying the optimum fused feature. The effectiveness of proposed method is verified by 222 pairs of mammograms from 222 women in Northeast China through complete experiment. The new method with fused feature model, genetic selection and ELM classifier can provide radiologists with a valuable opinion to improve the accuracy and efficiency of early-stage breast cancer detection and diagnosis.

**Acknowledgments** This research was partially supported by the National Natural Science Foundation of China under Grant Nos. 61402089, 61100022, 61472069, and the National High Technology Research and Development Plan (863 Plan) under Grant No. 2012AA02A606. The authors wish to express their sincere appreciation to Zhongzhou Chen, Shiya Zhang and Sheng Gu for revising and polishing.

## References

1. Jatoi I, Miller AB (2003) Why is breast-cancer mortality declining? Lancet Oncol 4(4):251–254
2. Kopans DB (1997) The most recent breast cancer screening controversy about whether mammographic screening benefits women at any age: nonsense and nonscience. Am J Roentgenol 180(1):21–26
3. Maggio CD (2004) State of the art of current modalities for the diagnosis of breast lesions. Phys Med Biol 31(1):856–869
4. Vyborny CJ, Giger ML (1994) Computer vision and artificial intelligence in mammography. Phys Med Biol 162(3):699–708
5. Hoffman KR, Gray JE (1999) In the next decade automated computer analysis will be an accepted sole method to separate normal from abnormal radiological images. Med Phys 26(1):1–4
6. Huang G-B, Chen L, Siew C-K (2006) Universal approximation using incremental constructive feedforward networks with random hidden nodes. IEEE Trans Neural Netw 17(4):879C–892C
7. Huang G-B, Ding X, Zhou H (2010) Optimization method based extreme learning machine for classification. Neurocomputing 74:155C–163C
8. Huang G-B, Zhu Q-Y, Siew C-K (2006) Extreme learning machine: theory and applications. Neurocomputing 70:489C–501C
9. Huang G-B, Zhou H, Ding X-J, Zhang R (2011) Extreme learning machine for regression and multiclass classification. IEEE Trans Syst Man Cybern 42(2):513–529
10. Huang G-B, Chen L (2007) Convex incremental extreme learning machine. Neurocomputing 70:3056–3062
11. Rong H-J, Ong Y-S, Tan A-H, Zhu Z-X (2008) A fast pruned-extreme learning machine for classification problem. Neurocomputing 72:359–366
12. Wei X-K, Li Y-H (2008) Linear programming minimum sphere set covering for extreme learning machines. Neurocomputing 71:570–575
13. Suresh S, Sundararajan N, Saratchandran P (2008) A sequential multi-category classifier using radial basis function networks. Neurocomputing 71:1345–1358
14. Huang G-B, Chen L (2008) Enhanced random search based incremental extreme learning machine. Neurocomputing 71:3460–3468
15. Huang G-B, Li M-B, Chen L, Siew C-K (2008) Incremental extreme learning machine with fully complex hidden nodes. Neurocomputing 71:576–583
16. Lim J-S (2008) Recursive DLS solution for extreme learning machine-based channel equalizer. Neurocomputing 71:592–599
17. Huang G-B, Wang D-H, Lan Y (2011) Extreme learning machines: a survey. Int J Mach Learn Cybern 2:107–122
18. Kegelmeyer WP Jr (1993) Evaluation of stellate lesion detection in a standard mammogram data set. Int J Pattern Recognit Artif Intell 7(6):1477–1492
19. Chan H-P, Wei D-T, Helvie MA, Sahiner B, AdIer DD, Goodsitt MM, Petrick N (1995) Computer-aided classification of mammographic masses and normal tissue : linear discriminant analysis in texture feature space. Phys Med Biol 40(5):857–876
20. Karssemeijer N, te Brake GM (1996) Detection of stellate distortions in mammograms. IEEE Trans Med Imaging 15(5):611–619
21. Sahiner B, Chan H-P, Petrick N, Helvie MA, Goodsitt MM (1998) Computerized characterization of masses on mammograms: the rubber band straightening transform and texture analysis. Med Phys 25(4):516–526
22. Mudigonda NR, Rangayyan RM, Leo Desautels JE (2000) Gradient and texture analysis for the classification of mammographic masses. IEEE Trans Med Imaging 19(10):1032–1043
23. Sahiner B, Chan H-P, Petrick N, Helvie MA, Hadjiiski LM (2001) Improvement of mammographic mass characterization using spiculation measures and morphological features. Med Phys 28(7):280–291
24. Shi J-Z, Sahiner B, Chan H-P, Ge J, Hadjiiski L, Helvie MA, Nees A, Wu Y-T, Wei J, Zhou C, Zhang Y-H, Cui J (2008) Characterization of mammographic masses based on level set segmentation with new image features and patient information. Med Phys 35(1):280–290
25. Engeland S, Timp S, Karssemeijer N (2006) Finding corresponding regions of interest in mediolateral oblique and craniocaudal mammographic views. Med Phys 33(9):3203–3212

26. Pu J-T, Zheng B, Leader JK, Gur D (2006) Multiview-based computer-aided detection scheme for breast masses. Med Phys 33(9):3135–3143
27. Engeland S, Karssemeijer N (2007) Combining two mammographic projections in a computer aided mass detection method. Med Phys 34(3):898–900
28. Samulski M, Karssemeijer N (2008) Matching mammographic regions in mediolateral oblique and cranio caudal views: a probabilistic approach. SPIE Med Imaging 6915(16):2444–2451
29. Pu J-T, Zheng B, Joseph KL, David G (2008) An ellipse-fitting based method for efficient registration of breast masses on two mammographic views. Med Phys 35(2):487–494
30. Chang Y-H, Good WF, Leader JK, Wang X-H, Zheng B, Hardesty LA, Hakim CM, Gur D (2003) Integrated density of a lesion: a quantitative, mammographically derived, invariable measure. Med Phys 30(7):1805–1811
31. Ojala T, Pietikainen M, Maenpaa T (2002) A generalized local binary pattern operator for multiresolution gray scale and rotation invariant texture classification. Pattern Anal Mach Intell 24(7):971–987
32. Tourassi GD, Harrawood B, Singh S, Lo JY, Floyd CE (2007) Evaluation of information-theoretic similarity measures for content-based retrieval and detection of masses in mammograms. Med Phys 34(1):140–150
33. Wang Z, Yu G, Kang Y, Zhao Y, Qu Q (2014) Breast tumor detection in digital mammography based on extreme learning machine. Neurocomputing 128(3):175–184
34. Wang Z-Q, Kang Y, Yu G, Zhao Y-J (2013) Breast tumor detection algorithm based on feature selection ELM. J Northeast Univ (Nat Sci) 34(6):792–796