

Accepted Manuscript

Microcalcification Diagnosis in Digital Mammography using Extreme Learning Machine Based on Hidden Markov Tree Model of Dual-Tree Complex Wavelet Transform

Kai Hu , Wei Yang , Xieping Gao

PII: S0957-4174(17)30390-1
DOI: [10.1016/j.eswa.2017.05.062](https://doi.org/10.1016/j.eswa.2017.05.062)
Reference: ESWA 11354



To appear in: *Expert Systems With Applications*

Received date: 1 December 2016
Revised date: 5 May 2017
Accepted date: 26 May 2017

Please cite this article as: Kai Hu , Wei Yang , Xieping Gao , Microcalcification Diagnosis in Digital Mammography using Extreme Learning Machine Based on Hidden Markov Tree Model of Dual-Tree Complex Wavelet Transform, *Expert Systems With Applications* (2017), doi: [10.1016/j.eswa.2017.05.062](https://doi.org/10.1016/j.eswa.2017.05.062)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Highlights

- We proposed a novel microcalcification diagnosis method in digital mammography.
- The combined HMT and wavelet features are used to describe the microcalcifications.
- The evaluation is performed on the Nijmegen, MIAS and DDSM datasets.
- The results show the effectiveness of the proposed method in accuracy and stability.

Microcalcification Diagnosis in Digital Mammography using Extreme Learning Machine Based on Hidden Markov Tree Model of Dual-Tree Complex Wavelet Transform

Kai Hu^{a,b}

^a*The MOE Key Laboratory of Intelligent Computing & Information Processing, Xiangtan University, Xiangtan 411105, China*

^b*College of Information Engineering, Xiangtan University, Xiangtan 411105, China
Email: kaihu@xtu.edu.cn*

Wei Yang^{a,b}

^a*The MOE Key Laboratory of Intelligent Computing & Information Processing, Xiangtan University, Xiangtan 411105, China*

^b*College of Information Engineering, Xiangtan University, Xiangtan 411105, China
Email: 16043395@qq.com*

Xieping Gao^{a,b}

Corresponding author

^a*The MOE Key Laboratory of Intelligent Computing & Information Processing, Xiangtan University, Xiangtan 411105, China*

^b*College of Information Engineering, Xiangtan University, Xiangtan 411105, China
Email: xpgao@xtu.edu.cn
Tel: +86-731-58293676
Fax: +86-731-58292201*

Abstract

Diagnosis of benign and malignant microcalcifications in digital mammography using Computer-aided Diagnosis (CAD) system is critical for the early diagnosis of breast cancer. Wavelet transform based diagnosis methods are effective to accomplish this task, but limited by representing the correlation within each wavelet scale, these methods neglect the correlation between wavelet scales. In this paper, we apply the hidden Markov tree model of dual-tree complex wavelet transform (DTCWT-HMT) for microcalcification diagnosis in digital mammography. DTCWT-HMT can effectively capture the correlation between different wavelet coefficients and model the statistical dependencies and non-Gaussian statistics of real signals, is used to characterize microcalcifications for the diagnosis of benign and malignant cases. The combined features which consist of the DTCWT-HMT features and the DTCWT features are optimized by genetic algorithm (GA). Extreme learning machine (ELM), an efficient learning theory is employed as the classifier to diagnose the benign and malignant microcalcifications. The validity of the proposed method is evaluated on the Nijmegen, MIAS and DDSM datasets using area under curve (AUC) of receiver operating characteristic (ROC). The AUC values of 0.9856, 0.9941 and 0.9168 of the proposed method are achieved on Nijmegen, MIAS and DDSM, respectively. We compare the proposed method with state-of-the-art diagnosis methods, and the experimental results show the effectiveness of the proposed method for the diagnosis of the benign and malignant microcalcifications in mammograms in terms of the accuracy and stability.

Keywords: Microcalcification diagnosis; digital mammography; dual-tree complex wavelet transform; hidden Markov tree model; extreme machine learning; feature extraction.

1. Introduction

Reported by the American Cancer Society (2008), breast cancer is the second most common cancer in women all over the world. Early detection and diagnosis is essential for decreasing the death rate caused by breast cancer. Recently, digital mammography becomes one of the most effective techniques for early detection and diagnosis of breast cancer (Alayhoglu, & Aghdasi, 1999; Ganesan et al., 2013; Lee, & Chen, 2015). Microcalcification is as one of the important signs of breast cancer. The earlier the microcalcifications are detected, the greater the chance of cure can be provided. Fig. 1 shows the examples of benign and malignant cases of microcalcifications in mammograms. In clinical practice, correct detection of microcalcifications is very difficult (Ciecholowski, 2017). Especially it is difficult and time consuming for breast radiologists to distinguish malignant from benign microcalcifications (Chen et al., 2015). In order to improve the diagnostic accuracy and make the diagnostic process easier for breast radiologists, computer-aided diagnosis (CADx) is developed as the “second reader” which can provide useful information for the diagnosis of microcalcifications in mammograms.

In the past decades, there are a series of breast cancer CADx algorithms have been developed in the literature (Beura, Majhi, & Dash, 2015; Buciu, & Gacsadi, 2011; Chen et al., 2012, 2015; Choi et al., 2016; Ciecholowski, 2017; Crouse, Nowak, & Baraniuk, 1998; Dhawan et al., 1996; Diaz-Huerta et al., 2014; Görgel et al., 2009, 2013; Hamid, Farshid, & Siamak, 2004; Hu, Gao, & Li, 2011; Jen, & Yu, 2015; Jiang, Zhang, & Li, 2015; Krishnan et al., 2010; Mousa et al., 2005; Nascimento et al., 2013; Orchard, & Ramchandran, 1994; Rao, & Subramanyam, 2008; Strange, et al., 2014; Wajid, & Hussain, 2015; Zhang, & Gao, 2007). Overall reviewing the literature, wavelet transform has been widely used in this area, because it

can analyze signals in both time and frequency domain and can provide multiscale information of signals.

Dhawan et al. (1996) employed wavelet features and histogram statistics to describe the local textural information and the global textural information of microcalcifications respectively. Hamid, Farshid, and Siamak (2004) presented a microcalcification classification algorithm in mammogram. In their research they demonstrated that the classification result based on multiwavelet features performs better than those of based on single wavelet features. Zhang and Gao (2007) considered the translation-invariant wavelet transform, the redundancy provided by which can represent the texture of microcalcification better. Görgel et al. (2009) proposed a support vector machine (SVM) method in wavelet domain for the classification of mammographic masses. In 2013 they proposed another method (Görgel et al., 2013) in which the local seed region growing (LSRG) algorithm is employed for the detection of the region of interests (ROIs), the spherical wavelet transform is used for feature description and SVM is applied as classifier. Buciu and Gacsadi (2011) used Gabor wavelets and directional features for the classification of mammograms. After feature extraction, principal component analysis (PCA) is used for reducing the dimension of features and SVM is employed for classifying the data. Beura, Majhi, and Dash (2015) used the discrete wavelet transform (2D-DWT) and gray-level co-occurrence matrix (GLCM) for mammogram classification. After two dimensional DWT (2D-DWT) for each ROI of a mammogram, the feature matrix can be obtained on detailed coefficients through GLCM.

Although wavelet based CAD methods have been used in this area, most of which are restricted to represent the correlation within each wavelet scale and little attention has been put on the correlation between wavelet scales. Actually, the properties of wavelet coefficients such as residual dependency structures tend to propagate across scales (Orchard, & Ramchandran, 1994). In 1998, Crouse, Nowak, and

Baraniuk (1998) proposed the hidden Markov tree (HMT) model for modeling the statistical dependencies and non-Gaussian statistics of real signals. HMT model satisfies the probability density distribution of single wavelet coefficients and can capture the correlation between different wavelet coefficients (Rao, & Subramanyam, 2008). Herein, we utilize the HMT model in wavelet domain to extract texture feature of microcalcifications. For the wavelet basis, dual-tree complex wavelet transform (DTCWT) is used. That is because DTCWT can provide approximate shift invariance and can improve the angular resolution using a dual tree of wavelet filters to obtain their real and imaginary parts of coefficients (Kingsbury, 1998, 1999, 2001), which are useful for the representation of microcalcifications. As analyzed above, the HMT model and DTCWT for feature extraction of microcalcification have their respective advantages. The result of microcalcification diagnosis using the combined HMT model and DTCWT features is better than using either the HMT model features or DTCWT features alone. In this paper, we apply a complex HMT model of DTCWT (DTCWT-HMT) to characterize microcalcification clusters in digital mammography for the diagnosis of benign and malignant cases. The combined features come from two parts: DTCWT-HMT based features and DTCWT based features. Further, we adopt the genetic algorithm (GA) (Hamid, Farshid, & Siamak, 2004; Lin, Prasad, & Saxena, 2015) to optimize the various features to generate effective feature vector for improving the classification accuracy.

Next, a very important step is to use an effective classifier to distinguish the ROIs of mammograms into benign and malignant microcalcifications. Extreme learning machine (ELM) (Huang, Chen, & Siew, 2006; Huang, Zhu, & Siew, 2004, 2006), a very simple and effective learning algorithm, shows a variety of benefits: the mathematical derivation is simple, the learning speed is fast, the nodes of hidden layer need not be tuned, etc. Therefore, ELM has been widely applied in solving regression, classification, clustering,

feature learning, compressive sensing theory and application problems (Huang, 2014; Li, Yang, & Burdet, 2016; Sun, Liu, Huang, & Zhang, 2016; Tang, Deng, & Huang, 2016; Tang, & Han, 2009), such as breast cancer detection and classification (Malar et al., 2012; Xie, Li, & Ma, 2016; Wang et al., 2014). In this paper, after feature extraction and selection, we apply ELM as the classifier to diagnose microcalcifications.

In this paper, a novel CAD method, which takes advantage of both DTCWT-HMT and DTCWT features, is developed to provide effective and comprehensive feature description for the diagnosis of microcalcifications in mammograms. Genetic algorithm and ELM are employed in the proposed method for improving diagnostic accuracy and stability. The results on the Nijmegen, MIAS and DDSM datasets show the competitiveness of the proposed method when compared to other diagnosis methods. The rest of this paper is organized as follows. In Section 2, dual-tree complex wavelet transform, the hidden Markov tree model of DTCWT, and ELM are briefly reviewed. In Section 3, the proposed diagnosis method is described in detail. The experiments and discussion on three digital mammography datasets are presented in Section 4. Finally, concluding remarks and future work are given in Section 5.

2. Theoretical background

2.1. DTCWT

As presented by Kingsbury (1998, 1999, 2001), the DWT for some signal analysis applications has been hampered by lack of shift invariance. One possible method is to use the wavelet transform without decimation or use translation invariant wavelet transform for instead, but this will lead to high computation and high redundancy, especially in multiple dimensions (Kingsbury, 1998).

The coefficients of DTCWT are obtained using a dual tree of wavelet filters. As declared by Kingsbury (1998), DTCWT can provide approximate shift invariance with a limited redundancy ($2^m : 1$ for

m-dimensional signals) and improve angular resolution (including six oriented subbands, $\pm 15^\circ$, $\pm 45^\circ$, $\pm 75^\circ$ compared to DWT). Moreover, DTCWT requires much less computation because of the lower redundancy and its simpler separable filters (Kingsbury, 1998). The 2-D DTCWT is shown in Fig. 2. The new Q-shift dual-tree (Selesnick, Baraniuk, & Kingsbury, 2005) is adopted in our experiments.

2.2. The Hidden Markov tree model of DTCWT

The HMT model for real wavelet transform has been successfully applied in pattern recognition (He, You, & Tang, 2008). In this paper, we will employ the HMT of DTCWT for the diagnosis of benign and malignant cases of microcalcifications. Conveniently, we noted it as DTCWT-HMT model.

In DTCWT-HMT, we associate with each complex wavelet coefficient t_i a hidden state g_i which can take value L or S depending on whether $|t_i|$ is large or small (Choi, Romberg, Baraniuk, & Kingsbury, 2000). Assume each t_i as a random vector (u_i, v_i) , t_i can be modeled by 2-state Gaussian mixture density as

$$f(t_i) = \sum_{a=S,L} p(g_i = a) f(t_i | g_i = a)$$

where $p(g_i = a)$ is the probability mass function (PMF) of state variable g_i with value a and $f(t_i | g_i = a)$ is the Gaussian conditional probability density function which can be formulated using mean $\mu_{i,a}$ and variance $\sigma_{i,a}$ as

$$f(t_i | g_i = a) = \frac{1}{\sqrt{2\pi}\sigma_{i,a}} \exp\left[-\frac{(u_i - \mu_{i,a})^2 + (v_i - \mu_{i,a})^2}{2\sigma_{i,a}^2}\right]$$

since the wavelet coefficients of images are typically zero mean, we set $\mu_{i,a} = 0$.

As shown in Fig. 3, let links denote dependencies, the DTCWT-HMT model applies a Markov-dependency structure to the hidden states across scale (Crouse, Nowak, & Baraniuk, 1998). Compared to the real HMT, the DTCWT-HMT is a complex HMT, which has six subband trees instead of three. Let $\rho(i)$ denotes the index number of the parent of node i , $\epsilon_{i,\rho(i)}^{a,b} = p(t_i = a | t_{\rho(i)} = b)$ denotes the

conditional probability that a child t_i has hidden state a when its parent $t_{\rho(i)}$ has state b . Actually, this state transition probability models the persistence property of coefficients across scale (Crouse, Nowak, & Baraniuk, 1998). Based on the discussion above, the DTCWT-HMT model therefore can be defined by $\theta = \{p(t_i = a), \mu_{i,a}, \sigma_{i,a}^2, \varepsilon_{i,\rho(i)}^{a,b}\}$, which can be trained to obtain wavelet-domain features of the images using the Expectation Maximization (EM) estimation. That is, let t denote the complex wavelet coefficients of the image, the model $f(t|\theta)$ can approximate the joint PDF of t given the training data. See Crouse, Nowak, and Baraniuk, 1998; Choi, Romberg, Baraniuk, and Kingsbury, 2000 for more details.

2.3. Extreme learning machine

Recently, ELM developed by Huang, Zhu, and Siew (2004, 2006) has become one of the most popular machine learning theories. Unlike other traditional learning algorithms, ELM has only one hidden layer and the nodes of hidden layer need not be tuned. For S arbitrary distinct samples $(\mathbf{y}_i, \mathbf{l}_i)$, where $\mathbf{y}_i = [y_{i1}, y_{i2}, \dots, y_{in}]^T \in \mathbf{R}^n$ and $\mathbf{l}_i = [l_{i1}, l_{i2}, \dots, l_{im}]^T \in \mathbf{R}^m$, the output function of the SLFNs with activation function $\tau(y)$ and N hidden nodes can be modeled as the following equations (Huang, Zhu, & Siew, 2006)

$$f_N(\mathbf{y}_j) = \sum_{i=1}^N \alpha_i \tau(\mathbf{r}_i \cdot \mathbf{y}_j + \delta_i) = \mathbf{z}_j, \quad j = 1, \dots, S,$$

where $\mathbf{r}_i = [r_{i1}, r_{i2}, \dots, r_{in}]^T$ is the weights to express the relationships between the i th hidden node and the input nodes, $\alpha_i = [\alpha_{i1}, \alpha_{i2}, \dots, \alpha_{in}]^T$ is the weights to express the relationships between the i th hidden node and the output nodes, δ_i is the threshold of the i th hidden node, and $\mathbf{z}_j = [z_{j1}, z_{j2}, \dots, z_{jn}]^T$ is the output vector of the neural network. As presented by Huang, Zhu, and Siew (2006) the standard SLFNs can approximate the S samples with zero error given activation function $\tau(y)$ and the N hidden nodes, that means

$$\sum_{j=1}^S \|\mathbf{z}_j - \mathbf{l}_j\| = 0$$

and there exist α_i , \mathbf{r}_i and δ_i satisfy

$$\sum_{i=1}^N \alpha_i \tau(\mathbf{r}_i \cdot \mathbf{y}_j + \delta_i) = \mathbf{l}_j, \quad j = 1, \dots, S.$$

which can be rewritten as

$$\mathbf{H}\alpha = \mathbf{L} \quad (1)$$

where

$$\begin{aligned} & \mathbf{H}(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N, \delta_1, \delta_2, \dots, \delta_N, \mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_S) \\ &= \begin{bmatrix} \tau(\mathbf{r}_1 \cdot \mathbf{y}_1 + \delta_1) & \tau(\mathbf{r}_2 \cdot \mathbf{y}_1 + \delta_2) & \cdots & \tau(\mathbf{r}_N \cdot \mathbf{y}_1 + \delta_N) \\ \tau(\mathbf{r}_1 \cdot \mathbf{y}_2 + \delta_1) & \tau(\mathbf{r}_2 \cdot \mathbf{y}_2 + \delta_2) & \cdots & \tau(\mathbf{r}_N \cdot \mathbf{y}_2 + \delta_N) \\ \vdots & \vdots & \ddots & \vdots \\ \tau(\mathbf{r}_1 \cdot \mathbf{y}_S + \delta_1) & \tau(\mathbf{r}_2 \cdot \mathbf{y}_S + \delta_2) & \cdots & \tau(\mathbf{r}_N \cdot \mathbf{y}_S + \delta_N) \end{bmatrix}_{S \times N}, \quad \alpha = \begin{bmatrix} \alpha_1^T \\ \alpha_2^T \\ \vdots \\ \alpha_N^T \end{bmatrix}_{N \times m}, \text{ and } \mathbf{L} = \begin{bmatrix} \mathbf{l}_1^T \\ \mathbf{l}_2^T \\ \vdots \\ \mathbf{l}_S^T \end{bmatrix}_{S \times m}. \end{aligned}$$

\mathbf{H} represents the output of hidden layer in SLFNs. As presented by Huang, Zhu, and Siew (2004, 2006), above linear system can be solved through the formulation as bellow

$$\|\mathbf{H}\hat{\alpha} - \mathbf{L}\| = \min_{\alpha} \|\mathbf{H}\alpha - \mathbf{L}\|$$

here

$$\hat{\alpha} = \mathbf{H}^\dagger \mathbf{L} \quad (2)$$

represents the smallest norm least squares solution of $\mathbf{H}\alpha = \mathbf{L}$ and \mathbf{H}^\dagger denotes the Moore-Penrose generalized inverse of matrix \mathbf{H} .

3. Proposed methodology

3.1. Feature extraction based on DTCWT-HMT model and DTCWT

The most important and difficult factor in pattern recognition is feature extraction which can directly affect the diagnosis results. Herein, we combine the DTCWT-HMT features and DTCWT features to generate better diagnosis results.

3.1.1 ROI extraction

The aim of microcalcification diagnosis is to classify the area of abnormality. Therefore, we should extract ROI which contains microcalcification clusters at first. General method is to cut out ROI with same size. However, the real size of each ROI is different, which can range from several pixels to hundreds of pixels, therefore we get ROI according to its own position and size. Being convenient for wavelet transform, we make the length of border to be the integral power of two.

3.1.2 Feature extraction

Although shape features have been widely used for microcalcification diagnosis by researchers, the results of these methods depend on the precision of their segment algorithms, because the shape features are extracted from individual microcalcification which must be segmented from ROIs. Limited by the tiny size of microcalcification, the weak comparison of gray-level in ROI and the noise in mammogram, segment algorithm can hardly be precise enough. As a result, texture features are used in the presented diagnosis method. To capture the correlation and dependency between complex wavelet coefficients better, we combine the DTCWT-HMT based features and DTCWT based wavelet features.

The original intention of using DTCWT-HMT model to characterize the features of microcalcifications is to train the DTCWT-HMT model for the selected ROIs, and then employ the trained DTCWT-HMT parameter set $\theta = \{p(g_1 = a), \mu_{i,a}, \sigma_{i,a}^2, \varepsilon_{i,\rho(i)}^{a,b}\}$ as the features for describing each ROI. The parameter set is listed as follows.

(1) $p(g_1 = a)$, the probability mass function of state value of the root node 1.

(2) $\mu_{i,a}, \sigma_{i,a}^2$, the mean and variance of Gaussian probability density function of complex wavelet coefficient t_i given state $g_i = a$.

(3) $\varepsilon_{i,\rho(i)}^{a,b}$, the *parent* \rightarrow *children* link between hidden states.

The DTCWT-HMT model used in our experiment has six subband trees. Because $\mu_{i,m}$ is constrained to be zero and the state m may be S or L, we extract 7 parameters as features from each subband tree.

Besides the DTCWT-HMT based features, we also extract wavelet based features. In the experiment, for each ROI of mammograms, a DTCWT with four levels of decomposition is used, and therefore it generates 26 sub-images for each ROI. Six statistics including energy, entropy and 1-4 order central moments (Mini, & Thomas, 2003) are calculated in each $N \times N$ sub-image using the following equations

$$\text{Energy} = \frac{\sum_i \sum_j p_{ij}^2}{N^2},$$

$$\text{Entropy} = -\frac{\sum_i \sum_j [\frac{p_{ij}^2}{\text{norm}^2}] \log_2 [\frac{p_{ij}^2}{\text{norm}^2}]}{\log_2^{N^2}},$$

where p_{ij} denotes the ij th pixel value and $norm^2 = \sum_i \sum_j p_{ij}^2$.

$$M = \frac{1}{Z} \sum_{d=1}^D f_d n_d,$$

$$V = \frac{1}{Z} \sum_{d=1}^D (f_d - u)^2 n_d,$$

$$Sk = \frac{1}{Z \times \sigma^2} \sum_{d=1}^D (f_d - u)^3 n_d,$$

$$ku = \frac{1}{Z \times \sigma^4} \sum_{d=1}^D (f_d - u)^4 n_d - 3,$$

where f_d , D , Z and n_d denote the d th gray level, the total number of gray levels, the number of pixels of image, and the number of pixels with gray level f_d , respectively.

Finally, the proposed feature extraction methods above result in a total of 198 texture features including 42 features come from DTCWT-HMT model and 156 features come from the detail sub-images of DTCWT.

3.2. Feature selection using genetic algorithm

In order to reduce the feature dimensions and computational complexity as well as improve the diagnosis accuracy, we use genetic algorithm (Hamid, Farshid, & Siamak, 2004; Lin, Prasad, & Saxena, 2015) for feature selection and optimization of the extracted microcalcification features. The area under curve (AUC) of receiver operating characteristic (ROC) is employed as the fitness function in genetic algorithm for optimization. The genetic algorithm code package used in the experiment is provided in Matlab 8.1. Table 1 shows the parameter settings of genetic algorithm.

3.3. Diagnosis

Microcalcification diagnosis is a two-category classifying problem virtually, which aims to classify microcalcification clusters into benign and malignant cases. We use ELM as the classifier for the diagnosis of microcalcifications. The leave-one-out cross validation method is employed in our method for training and testing. For S samples, one is set as the test sample and the remaining $S-1$ samples are used for the training samples. The process is repeated until each sample is set as the test sample for classification.

3.3.1. ELM for training

Let \mathbf{FM} denotes the feature matrix generated by the proposed feature extraction algorithm on the microcalcification ROI images. Before the process of microcalcification classification training, the feature matrix \mathbf{FM} is normalized to the real number which ranges from 0 to 1. The ELM training is shown in Algorithm 1. Firstly, for each hidden node, randomly assign input weights \mathbf{r}_i and biases δ_i . Secondly, compute \mathbf{H} according to \mathbf{r}_i and δ_i . Finally, compute the output weight α_i with the usage of matrix \mathbf{H} and the ground truth label set \mathbf{L} of the microcalcification ROI images to finish the training.

Algorithm 1 The ELM training

Input:

The feature matrix, \mathbf{FM} ;
 The ground truth label set, \mathbf{L} ;
 The number of hidden nodes, N ;
 The activation function, $\tau(y)$;

Output:

The parameters of ELM, $\mathbf{r}_i, \delta_i, \alpha_i, i=1,2,\dots,N$.
Step 1: Randomly assign input weights \mathbf{r}_i and δ_i ;
Step 2: Compute the output matrix \mathbf{H} of hidden layer according to \mathbf{FM} , \mathbf{r}_i and δ_i ;
Step 3: Compute the output weight $\alpha_i, i=1,2,\dots,N$ according to \mathbf{H} and \mathbf{L} using (2);
return The parameters $(\mathbf{r}_i, \delta_i, \alpha_i, i=1,2,\dots,N)$.

Algorithm 2 The ELM testing

Input:

The feature matrix, \mathbf{FM} ;
 The ground truth label set, \mathbf{L} ;
 The number of hidden nodes, N ;
 The activation function, $\tau(y)$;
 The ELM parameters, $(\mathbf{r}_i, \delta_i, \alpha_i, i=1,2,\dots,N)$;

Output:

The results of diagnosis by ELM, $\tilde{\mathbf{L}}$.
Step 1: Compute the output matrix \mathbf{H} of hidden layer according to \mathbf{FM} , \mathbf{r}_i and δ_i ;
Step 2: Obtain the diagnosis results $\tilde{\mathbf{L}}$ according to \mathbf{H} and $\alpha_i, i=1,2,\dots,N$ using (1);
return $\tilde{\mathbf{L}}$.

3.3.2. ELM for testing

The ELM parameters are obtained after the training procedure, which can be used for testing. The procedure of the ELM testing is presented in Algorithm 2. For the testing samples, the first step is to calculate H according to the r_i and δ_i , and then the diagnosis results can be obtained based on the H and α_i . Here, the leave-one-out cross validation method on a small dataset has smaller testing error, since it provides maximum size training samples (Kramer, & Aghdasi, 1999). The result in experimental section has demonstrated its stability.

3.4. The procedures of the proposed diagnosis method

As shown in Fig. 4, an outline of the proposed diagnosis method can be summarized as the following procedures:

Step 1 Reprocessing: Noise equalization (Karssemeijer, 1993).

Step 2 Feature extraction based on DTCWT-HMT model and DTCWT from all ROIs.

Step 3 Normalization of the feature matrix.

Step 4 Feature selection using genetic algorithm.

Step 5 Diagnose all ROI images into benign and malignant cases of microcalcifications using ELM.

4. Experimental results and discussions

4.1. Mammogram dataset

The three normative mammogram datasets used in our experiments are the Nijmegen (Karssemeijer, 1993) in the University Hospital Nijmegen from Netherlands, the mammographic image analysis society (MIAS) (Mini, & Thomas, 2003) from Britain, and the Digital Database for Screening Mammography (DDSM) (Heath et al., 2001) from United States. The Nijmegen dataset includes 40 mammograms which are come from 21 different patients. The size of these mammograms is 2048×2048 . In this research we have

extracted 103 ROIs including 29 benign and 74 malignant clusters. To be different, there are 322 mammograms in MIAS digitized by the spatial resolution of $50\mu m$, 20 of which are the cases of microcalcifications. We have extracted 26 ROIs with 12 benign and 14 malignant clusters. The DDSM dataset contains 2620 screening cases which are digitized by one of four different scanners: DBA M2100 ImageClear, Lumisys 200 Laser, Howtek 960, and Howtek MultiRad850. In this study, we have extracted 150 ROIs from the DDSM dataset including 82 benign and 68 malignant clusters.

4.2. Evaluation metric

In the experiment, we use the ROC analysis as evaluation metric for the diagnosis performance. In ROC curve, Y-axis denotes True positive fraction (TPF), or sensitivity, means that the correct classification rate of malignant ROIs. X-axis denotes the False positive fraction (FPF), which is also noted as (1-specificity). It defines the probability of the benign ROIs are misjudged as malignant ROIs. The TPF and FPF are, respectively, computed as follows

$$\begin{aligned} TPF &= \text{sensitivity} = \frac{TP}{TP + FN}, \\ FPF &= 1 - \text{specificity} = 1 - \frac{TN}{TN + FP} = \frac{FP}{TN + FP}, \end{aligned}$$

where, TN, TP, FN, and FP denote the numbers of true negatives, true positives, false negatives, and false positives, respectively. In the curve TPF and FPF are mutual restraint for each other, to improve the TPF will make lower FPF, and vice versa. AUC also known as A_z , which take values between 0 and 1, the higher the value of AUC, the better is the performance of the CAD system. In our method, the threshold is set for -1.4 to 1, which increases by 0.4 each time, and therefore the classification result can be draw a point of the ROC.

4.3. Experimental scheme and settings

4.3.1. Experiment on the Nijmegen and MIAS datasets

In order to evaluate the performance of the proposed method, especially for illustrating the influence of different feature sets and different classifiers on the classification performance, a series of experiments are conducted on the Nijmegen and MIAS datasets. In the experiments, we conduct the comparison based on

three feature sets, i.e. the experiment is performed on the wavelet features, the HMT model features, and the combination of both above, respectively. In addition, we use some other popular classifiers such as back-propagation (BP) neural network, SVM and K nearest neighbor (KNN) for the comparison to show the performance of ELM in classification efficiency.

(1) Different feature sets

The feature sets are divided into three parts: a) wavelet based features; b) HMT based features; and c) combined features based on wavelet and HMT model. For Part A, we compare the proposed diagnosis method on real wavelet transform based features and complex wavelet transform based features. In the experiments, we use wavelet, multiwavelet, and directional wavelet for real wavelet transform, and dual-tree complex wavelet for complex wavelet transform. We compare the proposed diagnosis method based on HMT features in wavelet domain for Part B. We extract HMT model parameters in both wavelet domain and dual-tree complex wavelet domain as features to classify ROI images into benign and malignant cases. We construct the HMT model of the chosen wavelets and extract the 1-4 levels model parameters as features on the corresponding three subband trees. While for DTCWT, the feature extraction is conducted on the six subband trees. In Part C, we compare the proposed diagnosis method on combined features between the HMT based features and wavelet based features. The wavelet functions for the comparison are presented in Table 2.

(2) Different classifiers

We use some typical classifiers such as BP, SVM and KNN for the comparison respectively. 1) BP. The selected neural network is a feedforward back-propagation three-layer perception with sigmoid nodes. Levenberg-Marquardt method is employed to train the three-layer network architecture, which

is considered to be a fast algorithm for training network (Lera, & Pinzolas, 2002). 2) SVM. SVM has recently received a widely considerable attention in pattern recognition (Papadopoulos, Fotiadis, & Likas, 2005; Zhang, O'Neill, Kong, & Grzegorczyk, 2008). In the experiment, we selected polynomial type of kernel function and the parameters of the kernel function were set with degree=1 and gamma=0.05. 3) KNN. This is a non-parametric method to label each testing sample's class based on its feature vector. The process of the classification is performed by calculating the similarity between a testing feature vector and each training feature vector and finding out the K nearest training cases. Therefore, the testing case is labeled by the class of the majority in these K cases belong. In our experiment, K is set to 7.

For SVM and KNN classifier, they are trained and tested using the leave-one-out cross validation. Limited by the great amount of computation, we train and test the BP neural network by two-fold cross validation method. We repeat the two-fold cross validation method for ten times and get the average result to avoid the error made by randomly select of testing and training samples. The threshold is set from -1 to k for KNN classifier to obtain the points of ROC curve. And for BP and SVM, the threshold is set as the same to our ELM-based diagnosis method.

For ELM diagnosis, we have evaluated the results by using different activation functions and various numbers of hidden nodes. A variety of investigations are conducted in order to optimize these two parameters. Table 2 shows the comparison of AUC values on different activation functions. The best performances are found with Triangular basis function on Nijmegen and Radial basis function on MIAS, respectively. Fig. 5 shows the performance comparison using various numbers of hidden nodes. The combinations that yield the best performance are found in Triangular basis function with 470 hidden nodes

on Nijmegen and Radial basis function with 400 hidden nodes on MIAS. The following experiments are conducted with above combinations.

4.3.2. Experiment on the DDSM dataset

To further evaluate the effectiveness of the proposed method, we conduct an experiment on the DDSM dataset. The extracted 150 ROIs from the DDSM dataset cover four breast tissue density categories (American College of Radiology, 2013) and five microcalcification types such as Amorphous distribution clustered, Pleomorphic distribution clustered, Pleomorphic distribution segmental, Punctate distribution segmental and Fine linear branching distribution clustered (Heath et al., 2001). The DDSM dataset is also used to evaluate the overall performance of the proposed method on different density categories and microcalcification types. Based on the results obtained on the Nijmegen and MIAS datasets, we choose Radial basis function as the activation function of ELM in this experiment, and the number of hidden nodes is set as 400.

4.4. Results and discussions

Table 3 shows the comparison of AUC measurements on the Nijmegen and MIAS datasets using different classifiers based on various feature sets. The ROC curves of the proposed diagnosis method using different classifiers are presented in Fig. 6, from which we can find that the AUC values of the proposed method is 0.9856 with a sensitivity of 92% and specificity of 93% on Nijmegen and is 0.9941 with a sensitivity of 100% and specificity of 92% on MIAS. Comparative analysis summarized in Table 3 show that the results of Part C based on HMT model features and wavelet features almost outperform those of Part B based on HMT model features or Part A based on wavelet features (see each column in Table 3 for

more details). In addition, in the comparison of AUC results by using a series of wavelet based features, we can find that the experimental results of 0.9632 for Nijmegen and 0.9792 for MIAS based on DTCWT are better than that of other wavelets. Furthermore, from the AUC values in Tables 3 we can conclude that ELM classifier perform better than other classifiers for all cases (see each row in Table 3 for more details). From the ROC curves presented in Fig. 6 it is also observed that ELM is the best of four. From what has been discussed above, the proposed diagnosis method using ELM based on hidden Markov tree model of dual-tree complex wavelet transform are perform better than those of other combination of features and classifiers.

Table 4-6 show the diagnosis results of the proposed method and other state-of-the-art methods. We can observe that the proposed method achieves the best results on the Nijmegen and MIAS dataset (see Tables 4 and 5). The increase in the AUC for the proposed diagnosis method compared to other state-of-the-art methods is 7.56 to 19.56% for the Nijmegen dataset, 4.37 to 19.41% for the MIAS dataset. Considering Malar's method (Malar et al., 2012) is employed for the diagnosis of normal and abnormal cases, the results of our diagnosis of benign and malignant cases can not be directly compared to Malar's, although our result 0.9941 is better than Malar's 0.98. From Table 6, it can be seen that although the result of our proposed method on the DDSM dataset is lower than Beura's method (Beura, Majhi, and Dash, 2015), it is still a good result and outperforms all other methods. As seen from Table 4-6, we can also observe that wavelet based methods outperform none-wavelet based methods. That is because the microcalcifications have high frequency in space domain and it is better to extract the high frequency information through wavelet decomposition, thus wavelet transform is an effective theory for the diagnosis of benign and malignant microcalcifications.

The results on the Nijmegen, MIAS and DDSM datasets have shown the diagnosis performance of the proposed method compared with state-of-the-art methods. Especially, the result on the DDSM dataset which contains different tissue density categories and microcalcification types shows the overall performance of the proposed method. From the results, we can observe that the proposed method achieves not only the high accuracy but also the high stability. The efficiency of the proposed method is derived from the following aspects. 1) The DTCWT can provide approximate shift invariance with a limited redundancy and can improve the angular resolution using a dual tree of wavelet filters. As a result, the DTCWT-based method is useful for the representation of microcalcifications. 2) HMT model can satisfy the probability density distribution of single wavelet coefficients and can capture the correlation between different wavelet coefficients. Therefore, it is suitable for modeling image feature information. 3) The result of microcalcification diagnosis using the combined DTCWT-HMT model and DTCWT features is better than using either the DTCWT-HMT model features or DTCWT features alone. Therefore, the proposed method, which takes advantage of both DTCWT-HMT and DTCWT features, is developed to provide effective and comprehensive feature description for the diagnosis of microcalcifications in mammograms. 4) We use genetic algorithm for selecting the best features using global search ability. The experimental results show that GA is an efficient optimization tool and is useful for the diagnosis of benign and malignant microcalcifications. 5) For the direct comparison, as shown in Table 3 and Fig. 6, ELM (as the classifier in the proposed method) achieves better performance than other popular classifiers.

The proposed method is implemented with MATLAB 8.1 on a PC (Intel i5-4590, 3.3-GHz CPU, and 8G RAM). In the experiments, genetic algorithm involves many generations and populations and therefore the GA-based feature selection process is quite slow. For other steps such as the feature extraction based on

DTCWT-HMT and DTCWT and the diagnosis using ELM, their running speed is fast. The complexity of the proposed method varies from dataset to dataset, since each dataset contains a different number of samples. The whole diagnosis system takes about 60 hours on Nijmegen, 14 hours on MIAS, and 90 hours on DDSM, respectively.

5. Conclusion and further research

In this paper, a novel microcalcification diagnosis system based on the combination of HMT based features and wavelet based features in dual-tree complex wavelet domain is developed. The DTCWT-HMT effectively models statistical distribution of wavelet coefficients and represents the correlation between wavelet coefficients well. After the combination of the DTCWT-HMT based features and DTCWT based features are extracted, they are improved by genetic algorithm and input to an effective and efficient classifier, ELM, for the diagnosis of benign and malignant microcalcification clusters. The experimental results using AUC evaluation verify that the proposed diagnosis method outperforms several state-of-the-art methods on the Nijmegen, MIAS, and DDSM datasets for its accuracy and stability. Further, comparisons of the proposed method using different feature sets have shown that the result using the combined DTCWT-HMT model and DTCWT features is better than using either the DTCWT-HMT model features or DTCWT features alone. Furthermore, comparisons of the proposed method using different classifiers have demonstrated that ELM provides the better diagnosis result than other classifiers.

Although the proposed method achieves a good result on the diagnosis of benign and malignant microcalcification clusters, it still can be improved. Further research can be carried out from the following aspects. The GA-based feature selection process is quite slow. We will investigate a fast and effective feature learning algorithm to create a discriminative feature set for characterizing microcalcifications in

mammograms, and therefore to reduce the computational complexity of the proposed method. The types of microcalcifications are much more than the five types included in our experiment. In our future work, we plan to evaluate the proposed method on bigger dataset which covers more types of microcalcifications. Furthermore, the ROI extraction for each mammogram is implemented based on the ground truth, which is provided by breast radiologists. We will investigate an automated method for the detection and segmentation of microcalcification regions to improve the applicability in practice.

Acknowledgements

The authors would like to thank the editor and reviewers for their insightful comments which have greatly helped to improve the quality of this paper. This work was supported by the National Natural Science Foundation of China under Grants 61172171 and 61401386 and by the Scientific Research Fund of Hunan Provincial Education Department under Grant 16C1545.

References

- Alayoglu, B. A., & Aghdasi, F. (1999). A wavelet neural network for the detection of microcalcification in multiscales enhanced digitized mammograms. In *AFRICON*.
- American Cancer Society. (2008). *Cancer Facts and Figures*. Atlanta: American Cancer Society Inc.
- American College of Radiology (2013). ACR BI-RADS—mammography. In: ACR BI-RADS atlas: breast imaging reporting and data system. 5th ed. Reston, Va: American College of Radiology.
- 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.

Buciu, I., & Gacsadi, A. (2011). Directional features for automatic tumor classification of mammogram images.

Biomedical Signal Processing and Control, 6(4), 370-378.

Chen, Z., Denton, E. R. E., & Zwigelaar, R. (2012). Classification of microcalcification clusters based on morphological topology analysis. In: A. Maidment, P. Bakic, S. Gavenonis (Eds.), *Breast Imaging, Lecture Notes in Computer Science*, Springer, Berlin, Heidelberg, 7361, 521-528.

Chen, Z., Strange, H., Oliver, A., Denton, E. R. E., Boggis, C., & Zwigelaar, R. (2015). Topological modeling and classification of mammographic microcalcification clusters. *IEEE Transactions on Biomedical Engineering*, 62(4), 1203-1214.

Choi, H., Romberg, J., Baraniuk, R., & Kingsbury, N. G. (2000). Hidden Markov tree modeling of complex wavelet transforms. In *Proceeding of the ICASSP*.

Choi, J. Y., Kim, D. H., Plataniotis, K. N., & Ro, Y. M. (2016). Classifier ensemble generation and selection with multiple feature representations for classification applications in computer-aided detection and diagnosis on mammography. *Expert Systems with Applications*, 46(15), 106-121.

Ciecholewski, M. (2017). Microcalcification segmentation from mammograms: A morphological approach. *Journal of Digital Imaging*, 30, 172-184.

Crouse, M. S., Nowak, R. D., & Baraniuk, R. G. (1998). Wavelet-based statistical signal processing using hidden Markov models. *IEEE Transactions on Signal Processing*, 46, 886-902.

Dhawan, A. P., Chitre, Y., Kaiser-Bonasso, C., & Moskowitz, M. (1996). Analysis of mammographic microcalcifications using grey-level image structure features. *IEEE Transactions on Medical Imaging*, 15(3), 246-259.

Diaz-Huerta, C.C., Felipe-Riveron, E. M., & Montaño-Zetina L. M. (2014). Quantitative analysis of morphological techniques for automatic classification of micro-calcifications in digitized mammograms. *Expert Systems with Applications*, 41(16), 7361-7369.

Ganesan, K., Acharya, U. R., Chua, C. K., Min, L. C., Abraham, K. T., & Ng, K.-H. (2013). Computer-aided breast cancer detection using mammograms: a review. *IEEE Reviews in Biomedical Engineering*, 6, 77-98.

Görgel, P., Sertbaş, A., Kilic, N., Ucan, O. N., & Osman, O. (2009). Mammographic mass classification using wavelet based support vector machine, *Istanbul University Journal of Electrical & Electronics Engineering*, 9(1), 867-875.

Görgel, P., Sertbas, A., & Ucan, O. N. (2013). Mammographical mass detection and classification using local seed region growing-spherical wavelet transform (lsrg-swt) hybrid scheme. *Computers in Biology and Medicine*, 43(6), 765-774.

Hamid, S. Z., Farshid, R. R., & Siamak, P. N. D. (2004). Comparison of multiwavelet, wavelet, Haralick, and shape features for microcalcification classification in mammograms. *Pattern Recognition*, 37(10), 1973-1986.

He, Z., You, X., & Tang, Y. (2008). Writer identification of Chinese handwriting documents using hidden Markov tree model. *Pattern Recognition*, 41(4), 1295-1307.

Heath, M., Bowyer, K., Kopans, D., Moore, R., & Kegelmeyer, W. P. (2001). The digital database for screening mammography. In *Proceedings of the Fifth International Workshop on Digital Mammography*, M.J. Yaffe, ed., 212-218, Medical Physics Publishing, ISBN 1-930524-00-5.

Hu, K., Gao, X., & Li, F. (2011). Detection of suspicious lesions by adaptive thresholding based on multiresolution analysis in mammograms. *IEEE Transactions on Instrumentation and Measurement*, 60(2), 462-472.

Huang, G.-B. (2014). An insight into extreme learning machines: Random neurons, random features and kernels. *Cognitive Computing*, 6(3), 376-390.

Huang, G.-B., Chen, L., & Siew, C.-K. (2006). Universal approximation using incremental constructive feedforward networks with random hidden nodes. *IEEE Transactions on Neural Network*, 17(4), 879-892.

Huang, G.-B., Zhu, Q.-Y., & Siew, C.-K. (2004). Extreme learning machine: A new learning scheme of feedforward neural networks. In *Proceeding of the IEEE International Joint Conference on Neural Networks*, 2, 985-990.

Huang, G.-B., Zhu, Q.-Y., & Siew, C.-K. (2006). Extreme learning machine: Theory and applications. *Neurocomputing*, 70(1), 489-501.

Jen, C., & Yu, S. S. (2015). Automatic detection of abnormal mammograms in mammographic images. *Expert Systems with Applications*, 42(6), 3048-3055.

Jiang, M., Zhang, S., Li, H., & Metaxas, D. N. (2015). Computer-aided diagnosis of mammographic masses using scalable image retrieval. *IEEE Transactions on Biomedical Engineering*, 62(2), 783-792.

Karssemeijer, N. (1993). Adaptive noise equalization and recognition of microcalcification clusters in mammograms. *International Journal of Pattern Recognition and Artificial Intelligence*, 7(6), 1357-1376.

Kingsbury, N. G. (1998). The dual-tree complex wavelet transform: a new technique for shift invariance and directional filters. *IEEE Digital Signal Processing Workshop (DSP 98)*, Bryce Canyon, No86.

Kingsbury, N. G. (1999). Image processing with complex wavelets. *Philosophical Transactions of the Royal Society of London A*, September 1999, on a Discussion Meeting on “Wavelets: the key to intermittent information?” London, 24-25.

Kingsbury, N. G. (2001). Complex wavelets for shift invariant analysis and filtering of signals. *Journal of Applied and Computational Harmonic Analysis*, 10, 234-253.

Kramer, D., & Aghdasi, F. (1999). Texture analysis techniques for the classification of microcalcifications in digitized mammograms. In *Proceeding of the 5th IEEE AFRICON Conference*, Cape Town, Africa, 395-400.

Krishnan, M. M. R., Banerjee, S., Chakraborty, C., Chakraborty, C., & Ray, A. K. (2010). Statistical analysis of mammographic features and its classification using support vector machine. *Expert Systems with Applications*, 37(1), 470-478.

Lee, H., & Chen, Y.-P. P. (2015). Image based computer aided diagnosis system for cancer detection. *Expert Systems with Applications*, 42(12), 5356-5365.

Lera, G., & Pinzolas, M. (2002). Neighborhood based Levenberg-Marquardt algorithm for neural network training. *IEEE Transactions on Neural Network*, 13(5), 1200-1203.

Li, Z., Yang, C., & Burdet, E. (2016). Guest editorial an overview of biomedical robotics and bio-mechatronics systems and applications. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 46(7), 869-874.

Lin, C.-T., Prasad, M., & Saxena, A. (2015). An improved polynomial neural network classifier using real-coded genetic algorithm. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 45(11), 1389-1401.

Malar, E., Kandaswamy, A., Chakravarthy, D., & Dhara, A. G. (2012). A novel approach for detection and classification of mammographic microcalcifications using wavelet analysis and extreme learning machine. *Computers in Biology and Medicine*, 42, 898-905.

Mini, M. G., & Thomas, T. (2003). A neural network method for mammogram analysis based on statistical features. In *TENCON*, 1489-1492.

Mousa, R., Munib, Q., & Moussa, A. (2005). Breast cancer diagnosis system based on wavelet analysis and fuzzy-neural. *Expert Systems with Applications*, 28, 713-723.

Nascimento, M. Z. do, Martins, A. S., Neves, L. A., Ramos, R. P., Flores, E. L., & Carrijo, G. A. (2013). Classification of masses in mammographic image using wavelet domain features and polynomial classifier. *Expert Systems with Applications*, 40(15), 6213-6221.

Orchard, M. T., & Ramchandran, K. (1994). An investigation of wavelet-based image coding using an entropy-constrained quantization framework. In *Proceedings of the Data Compression Conference*, Snowbird, UT, 341-350.

Papadopoulos, A., Fotiadis, D. I., & Likas, A. (2005). Characterization of clustered microcalcifications in digitized mammograms using neural networks and support vector machines. *Artificial Intelligence in Medicine*, 34(2), 141-150.

Rao, V. R., & Subramanyam, V. P. R. (2008). Rotation-invariant texture retrieval using wavelet-based hidden Markov trees. *Signal Processing*, 88(10), 2593-2598.

Ren, J. (2012). ANN vs. SVM: Which one performs better in classification of MCCs in mammogram imaging. *Knowledge-Based Systems*, 26, 144-153.

Selesnick, W., Baraniuk, R. G., & Kingsbury, N.G. (2005). The dual-tree complex wavelet transform. *IEEE Signal Processing Magazine*, 22(6), 123-151.

Strange, H., Chen, Z., Denton, E. R. E., & Zwigelaar, R. (2014). Modelling mammographic microcalcification clusters using persistent Mereotopology. *Pattern Recognition Letter*, 47, 157-163.

Sun, F., Liu, C., Huang, W., & Zhang, J. (2016). Object classification and grasp planning using visual and tactile sensing. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 46(7), 969-979.

Tang, J., Deng, C., & Huang, G.-B. (2016). Extreme learning machine for multilayer perceptron. *IEEE Transactions on Neural Network and Learning System*, 27(4), 809-821.

Tang, X., & Han, M. (2009). Partial Lanczos extreme learning machine for single-output regression problems. *Neurocomputing*, 72(13-15), 3066-3076.

Wajid, S. K., & Hussain, A. (2015). Local energy-based shape histogram feature extraction technique for breast cancer diagnosis. *Expert Systems with Applications*, 42(20), 6990-6999.

Wang, Z., Yu, G., Kang, Y., Zhao, Y., & Qu, Q. (2014). Breast tumor detection in digital mammography based on extreme learning machine. *Neurocomputing*, 128, 175-184.

Xie, W., Li, Y., & Ma, Y. (2016). Breast mass classification in digital mammography based on extreme learning machine. *Neurocomputing*, 173, 930-941.

Zhang, B., O'Neill, K., Kong, J. A., & Grzegorczyk, T. M. (2008). Support vector machine and neural network classification of metallic objects using coefficients of the spheroidal MQS response modes. *IEEE Transactions on Geoscience and Remote Sensing*, 46(1), 159-171.

Zhang, L., & Gao, X. (2007). Research on translation-invariant wavelet transform for classification in mammograms. In *the 3rd International Conference on Natural Computation*, 3, 571-575.

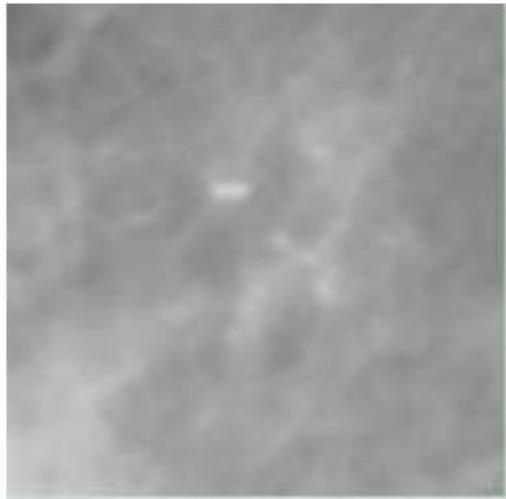
ACCEPTED MANUSCRIPT

Figure captions

Fig. 1. Examples of microcalcification clusters. a) Benign and b) Malignant.



a)



b)

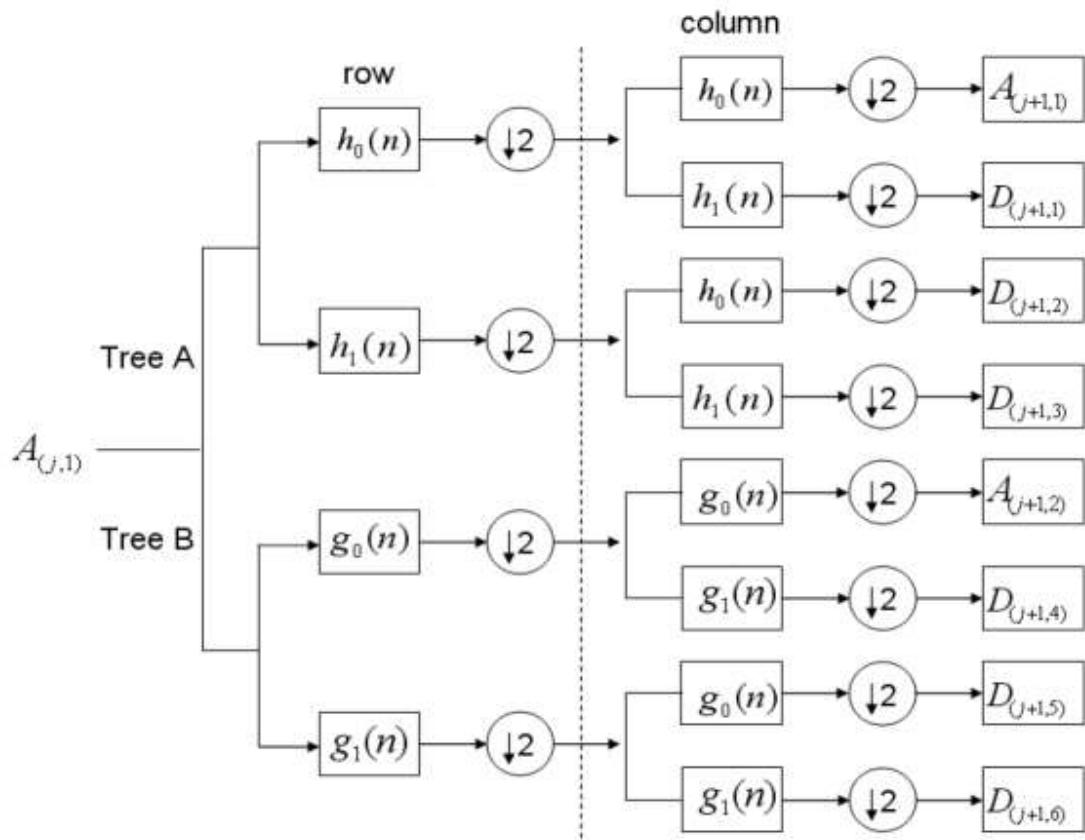
Fig. 2. The 2-D DTCWT.

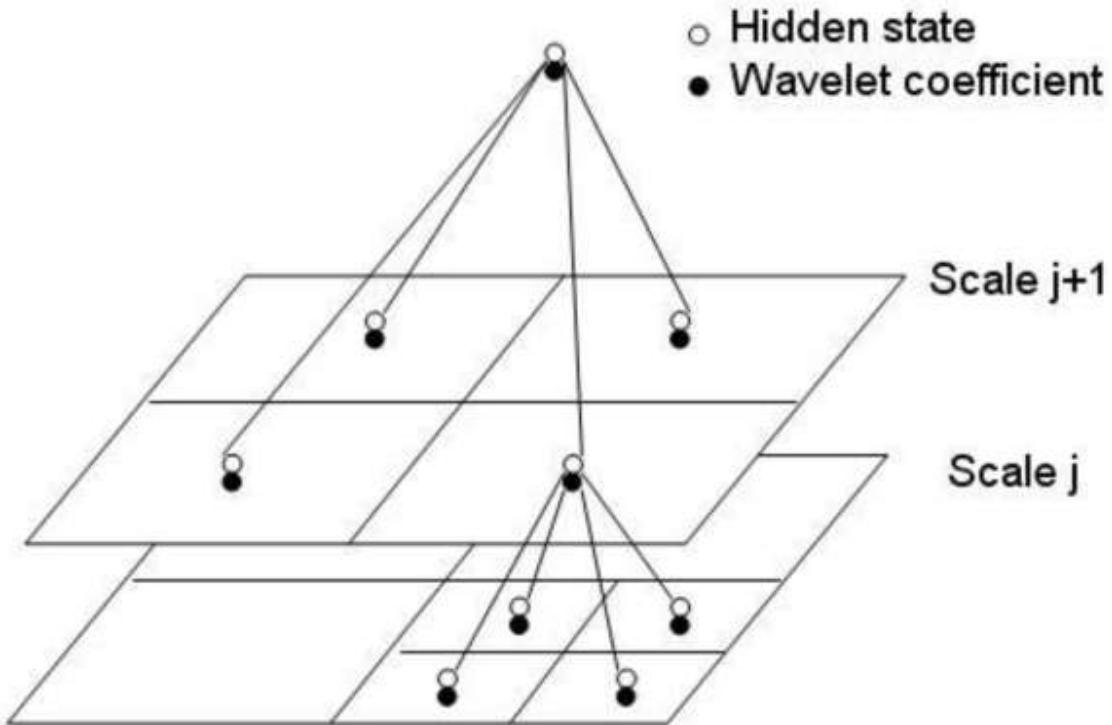
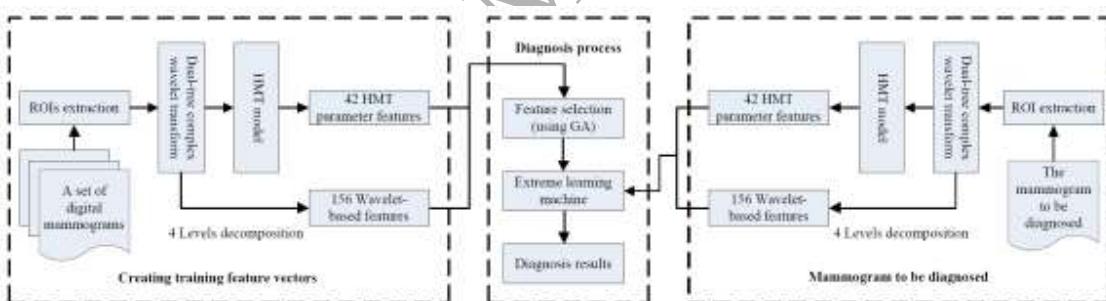
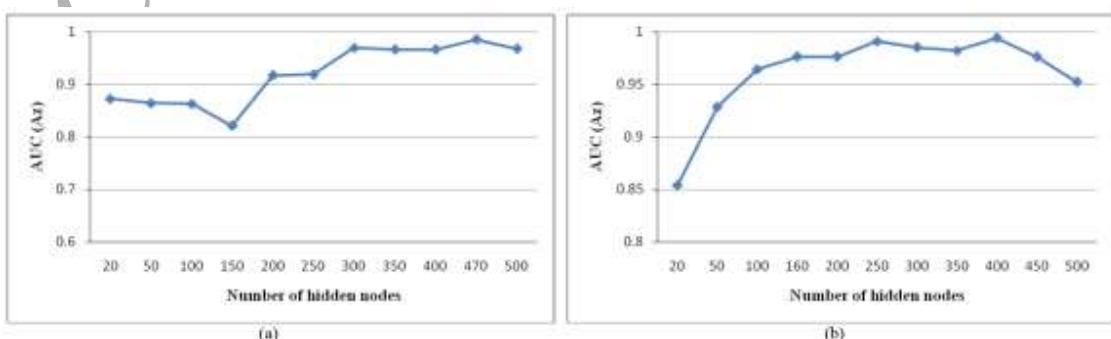
Fig. 3. The 2-D HMT model.**Fig. 4.** The proposed microcalcification diagnosis system.**Fig. 5.** Comparison of the AUC (A_z) values of the proposed method using various numbers of hidden nodes on (a) Nijmegen and (b) MIAS.

Fig. 6. ROC curves of the proposed method using different classifiers on (a) Nijmegen and (b) MIAS.

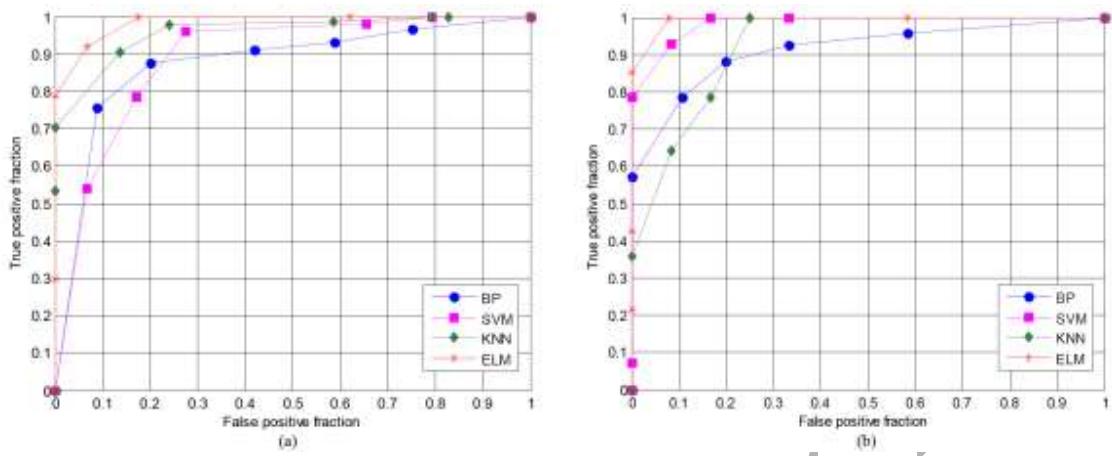


Table captions**Table1.** Parameter settings of genetic algorithm.

Selection operator	Stochastic uniform
Cross over operator	Two point
Probability of cross over	0.8
Mutation operator	Boundary mutation
Probability of mutation	0.03
Generations	200

Table 2. Performance comparison using different activation functions on the Nijmegen and MIAS datasets.

Activation function	AUC (Az)	
	Nijmegen	MIAS
Sigmoid	0.9651	0.9435
Sine	0.9644	0.8363
Hardlimit	0.9397	0.9345
Triangular basis function	0.9856	0.8542
Radial basis function	0.9811	0.9941

Table 3. Comparison of AUC (Az) measurements using different classifiers based on wavelet features.

Wavelet basis	Nijmegen				MIAS				
	BP	SVM	KNN	ELM	BP	SVM	KNN	ELM	
Diagnosis based on wavelet features									
Single wavelet	CDF97	0.9164	0.8989	0.9159	0.9569	0.8036	0.9286	0.9137	0.9732
	DB4	0.9090	0.9124	0.9331	0.9597	0.8869	0.9286	0.9167	0.9316
	SYM4	0.9087	0.9236	0.9425	0.9473	0.9048	0.9494	0.9001	0.9702
Multi wavelet	CL3	0.8728	0.9157	0.9138	0.9259	0.9464	0.9436	0.9302	0.9494
	GHM	0.8681	0.8616	0.8837	0.9247	0.8824	0.8962	0.9107	0.9583
	SA4	0.8688	0.8800	0.9238	0.9240	0.9083	0.9167	0.8899	0.9643
Directional wavelet	CDF97-Yue3	0.9203	0.8728	0.8996	0.9343	0.8244	0.9375	0.8905	0.9702
	DB4-Yue3	0.8406	0.8898	0.9199	0.9310	0.9375	0.9048	0.9048	0.9613
	SYM4-Yue3	0.8828	0.8812	0.9247	0.9266	0.9425	0.9521	0.8988	0.9673
Dual-tree complex wavelet	(af(1),af(2))	0.8583	0.9070	0.9469	0.9632	0.9167	0.9643	0.9406	0.9792
Diagnosis based on HMT model features									
Single wavelet	DB4	0.6700	0.7034	0.7849	0.8392	0.8631	0.8452	0.6250	0.9286

	SYM4	0.6456	0.6941	0.7493	0.8497	0.6845	0.6935	0.8214	0.8929
Dual-tree complex wavelet	(af(1),af(2))	0.7730	0.6897	0.7519	0.8637	0.9077	0.9048	0.8958	0.9405
Diagnosis based on HMT model features and wavelet features									
Dual-tree complex wavelet	(af(1),af(2))	0.8754	0.8886	0.9595	0.9856	0.9137	0.9851	0.9256	0.9941

Table 4. Comparison of the proposed method with other state-of-the-art methods on the Nijmegen dataset.

Method	Selected wavelet transform	Classifier	AUC (A_z)
Alayhoglu and Aghdasi (1999)	Wavelet packet transform using DB3 and DB10	NN	0.86
Hamid, Farshid, and Siamak (2004)	Multiwavelet transform using GHM, CL and SA4	KNN	0.89
Papadopoulos, Fotiadis, and Likas (2005)	-	SVM	0.79
Zhang and Gao (2007)	TI wavelet transform using DB4 and DB97	KNN	0.91
Ours	DTCWT-HMT model and DTCWT	KNN	0.9595
Ours	DTCWT-HMT model and DTCWT	ELM	0.9856

Table 5. Comparison of the proposed method with other state-of-the-art methods on the MIAS dataset.

Method	Selected wavelet transform	Classifier	AUC (A_z)
Papadopoulos, Fotiadis, and Likas (2005)	-	SVM	0.81
Chen, Denton, and Zwigelaar (2012)	-	KNN	0.93
Malar et al. (2012)	Harr wavelet transform	ELM	0.98
Ren (2012)	-	KNN	0.91
Strange et al. (2014)	-	Barcodes	0.80
Beura, Majhi, and Dash (2015)	Gray-level co-occurrence matrix and discrete wavelet transform	NN	0.9504
Chen et al. (2015)	-	KNN	0.96
Ours	DTCWT-HMT model and DTCWT	SVM	0.9851
Ours	DTCWT-HMT model and DTCWT	ELM	0.9941

Table 6. Comparison of the proposed method with other state-of-the-art methods on the DDSM dataset.

Method	Selected wavelet transform	Classifier	AUC (A_z)
Chen, Denton, and Zwigelaar (2012)	-	KNN	0.90
Ren (2012)	-	KNN	0.86
Strange et al. (2014)	-	Barcodes	0.82
Beura, Majhi, and Dash (2015)	Gray-level co-occurrence matrix and discrete wavelet transform	NN	0.9761
Chen et al. (2015)	-	KNN	0.90
Ours	DTCWT-HMT model and DTCWT	ELM	0.9168