A Hybrid Microcalcification Detection Algorithm In Digital Mammograms

Ping Zhang[a], Kwabena Agyepong[b]

[a] Department of Mathematics and Computer Science
[b] Department of Advanced Technologies
Alcorn State University, Mississippi 39096-7500, USA

Abstract—A novel hybrid wavelet-based fractal feature extraction method is proposed for the detection of microcalcification clusters (MCCs) in digital mammograms. The hybrid features consists of a set of the surrounding region dependence based features [11] and the newly proposed wavelet-based fractal features. A new fractal feature extraction scheme is given in this paper, which is based on the wavelet coefficients of a mammography image. Experiments demonstrated that the proposed hybrid features have the best convergence ability of artificial neural networks (ANNs) classifier compared to other two sets of features tested in the experiments. A good ratio of true positive fraction to false positive fraction (ROC curve) has been achieved. The proposed MCCs detection system provides an adequate framework for microcalcification detection in mammograms.

Keywords—pattern recognition, hybrid feature extraction, ANN classifier, calcification detection

I. Introduction

Breast cancer can be divided into three categories: microcalcifications, masses, and architectural distortions. An early sign of breast cancer is the presence of microcalcification clusters (MCCs) in the mammogram. MCCs are small in size and have low contrast that may be missed or misinterpreted by physicians and the task of eye-based mammography screening is tedious; therefore, a reliable and automatic computer-aided diagnosis system (CADx) could be very helpful to aid radiologists in detecting mammography lesions that may indicate the presence of breast cancer.

Microcalcifications are tiny deposits of calcium which appear as small bright spots on the mammogram. Microcalcifications are characterized by clusters, types, and distribution properties. Fig.1 shows two images of MCCs (a, b) and two images of normal mammogram (c, d).

Microclacification image analysis and detection is an extremely challenging task for the following three reasons: First of all, there is a large variability in the appearance of abnormalities. Likewise, abnormalities are often occluded or hidden in dense breast tissue. Perhaps most importantly, a CADx system for MCCs detection is used in serious human disease detection; hence, a need for near faultlessness is required.

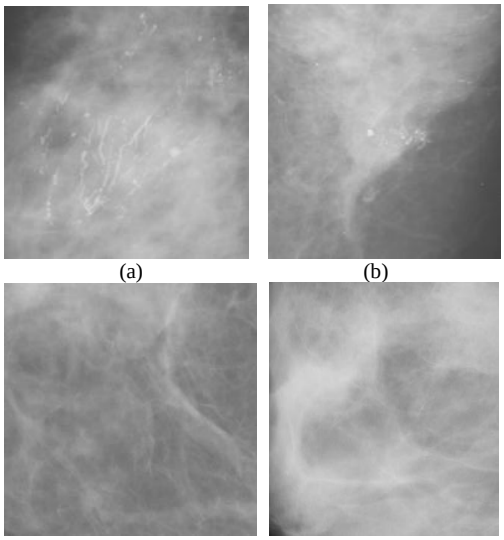

Fig. 1. Four mammograms: (a, b) calcification images; (c, d) normal images

Concerning calcification detection in the regions of interests (ROIs), many methods have been proposed. In the survey paper [1], mammogram enhancement and segmentation algorithms, mammographic features, classifiers and their performances were reviewed and compared. Remaining challenges were also discussed. In another paper [2], the detection performances of different classifiers, such as Support Vector Machines (SVMs), Kernel Fisher discriminant (KFD) classifier, Relevance Vector Machine (RVM), and committee machines (ensemble averaging and AdaBoost) were compared and the test results were reported. Neural networks have been used in many calcification systems. Two automatic microcalcification detection

Corresponding Author: Ping Zhang, Email Address: pzhang@alcorn.edu

systems were proposed based on the hybrid neural network classifier [3, 4]. SVMs have been used in the mammogram detection systems [5, 6]. Unsupervised detection of mammogram ROIs was introduced [7]. Segmentation of ROIs in mammogram using a topographic approach was introduced in the recent literature [8].

Feature extraction is one of the most important components in mammogram detection. A local feature extraction has been adopted in the literature more than a decade ago [9]. For example, the application of shape analysis to mammographic calcification was introduced [10]. A statistical textural feature for the detection of microcalcification was described in the reference [11]. Wavelets have been widely used in the feature extraction and segmentation in the mammogram detection [12-16]. Combining mathematical morphology and neural networks was also proposed in the literature [17] and multifractal analysis has been used in the medical image detection and classification [18-20].

In order to increase detection rate, a multiple expert system was given in the detection system [21]. A fuzzy logic was then introduced to detect calcification [22]. A microcalcifications detection algorithm by fitting a model to every location in the mammogram was proposed [23].

Artificial Neural Networks (ANNs) have been considered as efficient classifiers in many pattern recognition systems. Typically, an ANN accepts inputting features, which are computed from a specific region of interest, and provides an output as a characterization of the region [11].

In this paper, a microcalcification detection system with a novel hybrid feature extraction scheme is proposed. A system flowchart is drawn in Section II. Hybrid feature extraction method is presented in Section III, which concatenates the following feature sets: Surrounding Region Dependence Based (SRDM) feature set + Wavelet-based multifractal feature set. The specification of an ANN classifier is listed in Section IV. The experimental comparisons on calcification detection performance have been conducted on three ANN classifiers, which were trained by three sets of features. The testing results are reported in Section V. The conclusion is given in the last section.

II. Digital Mammogram Detection System

The schematic diagram of the proposed system is shown in Fig. 2. The system includes five components: ROI image input, image preprocessing, hybrid feature extraction, ANN classifier design and detection output.

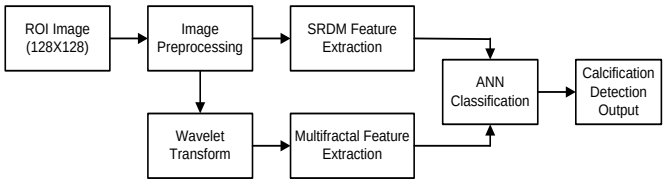

Fig. 2. Schematic diagram for microcalcification cluster detection

Image preprocessing includes the following operations on mammograms: image contrast enhancement, noise removal and histogram equalization in order to retain small spots (calcification feature) in the mammograms. For the detail image preprocessing algorithms, the interested readers can refer to image processing book [30].

Hybrid feature extraction method is divided into three steps: SRDM feature extraction, wavelet transform, and wavelet-based multifractal feature extraction.

III. Hybrid Feature Extraction

Feature extraction and classifier design are two of the most important steps in the design of a pattern recognition system. We will discuss three sets of features as follows:

3.1. Surrounding Region Dependence Based Method (SRDM)

A simple surrounding region dependence based method [11] is used as a first feature set. Firstly, the microcalfication area of a mammogram is divided into different overlapping 128x128 blocks (The image block size of 128x128 is considered based on mammogram resolution and computation simplicity); then the first feature set is extracted from each block. In order to systematically address this feature set, the diagram of the surrounding regions for current pixel *(x, y)* is shown in Fig. 3. Here, $\boldsymbol{R_1}$ and $\boldsymbol{R_2}$ are the inner surrounding region and the outer surrounding region, respectively. w_1, w_2, and w_3 denote the size of the three square windows.

In the following calculations, *n* is the maximal pixel number of R1 (inner region); *m* is the maximal pixel number of R2 (outer region).

$$\alpha(i,j) = \#\{(x,y) \mid c_{R_1}(x,y) = i, c_{R_2}(x,y) = j, (x,y) \in L_x \times L_y\}$$
$$c_{R_1}(x,y) = \#\{(k,l) \mid (k,l) \in R_1, [S(x,y) - S(k,l)] > q\}$$

$$c_{R_2}(x,y) = \#\{(k,l) \mid (k,l) \in R_2, [S(x,y) - S(k,l)] > q\}$$

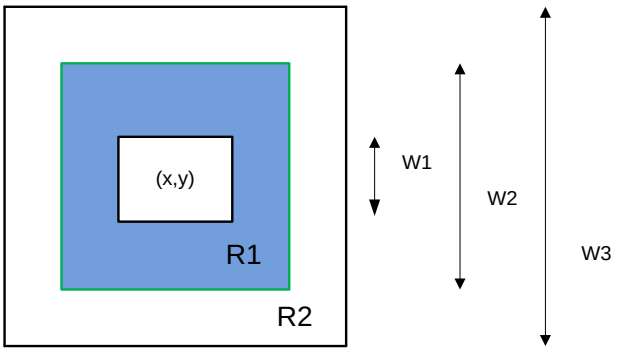

Fig. 3. Diagram of the surrounding regions

where *S(x, y)* is the image intensity of the current pixel (x, y), parameter *q* is a constant. # equates the number of pixels. *r(i,j)* is the reciprocal of the element, which is calculated as follows:

$$r(i,j) = \begin{cases} \dfrac{1}{\alpha(i,j)} & if\ \alpha(i,j) > 0 \\ 0 & otherwise \end{cases} \quad (1)$$

The following features are extracted:

Horizontal-weighted sum (HWS)

$$HWS = \frac{1}{N}\sum_{i=0}^{m}\sum_{j=0}^{n} j^2 r(i,j) \quad (2)$$

Vertical-weighted sum (VWS)

$$VWS = \frac{1}{N}\sum_{i=0}^{m}\sum_{j=0}^{n} i^2 r(i,j) \quad (3)$$

Diagonal-weighted sum (DWS)

$$DWS = \frac{1}{N}\sum_{k=0}^{m+n} k^2 \sum_{\substack{i=0 \\ i+j=k}}^{m}\sum_{j=0}^{n} r(i,j) \quad (4)$$

Grid-weighted sum (GWS)

$$GWS = \frac{1}{N}\sum_{i=0}^{m}\sum_{j=0}^{n} ijr(i,j) \quad (5)$$

where N is the total sum of elements in the surrounding region, which is defined as below:

$$N = \sum_{i=0}^{m}\sum_{j=0}^{n} \alpha(i,j) \quad (6)$$

Except for above mentioned four features, additional four new features are added in our system [24].

Variance sum (VS)

$$VS = \sum_{i=0}^{N_x-1}\sum_{j=0}^{N_y-1} (S(i,j) - \overline{S})^2 \quad (7)$$

Discrete Laplacian sum (DLS)

$$DLS = \sum_{i=0}^{N_x-1}\sum_{j=0}^{N_y-1} \frac{1}{4}\left(\frac{\partial^2 S(i,j)}{\partial i^2} + \frac{\partial^2 S(i,j)}{\partial j^2}\right) \quad (8)$$

Entropy (E)

$$E = -\sum_{i=0}^{N_x-1}\sum_{j=0}^{N_y-1} P(i,j)\log_2 P(i,j) \quad (9)$$

Mean Grey Level (MGL)

$$MGL = \frac{1}{N_x * N_y}\sum_{i=0}^{N_x-1}\sum_{j=0}^{N_y-1} S(i,j) \quad (10)$$

where N_x, N_y is the size of selected mammogram. In total, eight (8) features are extracted as texture features.

3.2. Wavelet Transform

The definition of a continuous wavelet transform is as follows[25]: for a continuous function *f(x)*, it is projected at each step *j* on the subset V_j, *(……⊂ V_{-1}⊂ V_0⊂ V_1⊂ V_2⊂...)*. The scalar project $c_{j,k}$ is defined by the dot product of *f(x)* with the scaling function $\phi(x)$, which is dilated and translated:

$$c_{j,k} =< f(x), \phi_{j,k}(x) >$$
$$\phi_{j,k}(x) = 2^{j/2}\phi(2^j x - k) \quad (11)$$

The difference between $c_{j+1,k}$ and $c_{j,k}$ is contained in the detailed component belonging to the space W_j, which is orthogonal to V_j. The following equations exist:

$$W_j \oplus V_j = V_{j+1}$$
$$V_j \cap W_j = \{0\},\ j \in Z \quad (12)$$

Suppose $\psi(x)$ is a wavelet function. The wavelet coefficients can be obtained by

$$w_{j,k} =< f(x), 2^{j/2}\psi(2^j x - k) > \quad (13)$$

Some relationships between $\phi(x)$ and $\psi(x)$ are listed below:

$$\tfrac{1}{2}\phi(\tfrac{x}{2}) = \sum_n h(n)\phi(x-n)$$
$$\tfrac{1}{2}\psi(\tfrac{x}{2}) = \sum_n g(n)\phi(x-n) \quad (14)$$

where *h(n)* and *g(n)* represent unit impulse functions of lowpass and highpass filters respectively, which are related to the scaling function $\phi(x)$ and the wavelet function $\psi(x)$.

According to the wavelet theory, a conventional two dimensional wavelet discrete transform (2D-DWT) can be regarded as being equivalent to filtering the input image with a bank of filters, whose impulse responses are all approximately given by scaled versions of a mother wavelet. The output of each level consists of four sub-images: *LL, LH, HL, HH* with 2:1 down-sampling.

Mathematically, we can express this recursive algorithm in the following equation.

$$\begin{aligned}\psi(x,y)_{LL} &= \phi(x)\phi(y)\\ \psi(x,y)_{LH} &= \phi(x)\psi(y)\\ \psi(x,y)_{HL} &= \psi(x)\phi(y)\\ \psi(x,y)_{HH} &= \psi(x)\psi(y)\end{aligned} \tag{15}$$

For example, The *LL* wavelet is the product of the low-pass function $\phi(x)$ along both the first dimension and second dimension; The *LH* wavelet is the product of the low-pass function $\phi(x)$ along the first dimension and the high-pass function $\psi(y)$ along the second dimension.

If the wavelet filters are real, then Mallat's dyadic wavelet decomposition fast algorithm [25] can be used. However, 2D-DWT has the following drawbacks: lack of shift invariance and poor directional selectivity.

3.3. Wavelet-based Multifractal Feature

The fractal application to image classification and recognition is receiving a lot of attention. The term fractal was coined by Mandelbrot in 1975 to describe the irregular structure of many natural objects and phenomena. Central to fractal geometry is the concept of self-similarity. Considering a bounded set ***R*** in Euclidean *n*-space, the set is said to be self-similar when ***R*** is the union of ***Nr*** distinct (non-overlapping) copies of itself, each of which has been scaled down by a ratio r<1 in all coordinates. The similarity dimension *Ds* is given by

$$N_r \cdot r^{Ds} = 1, \qquad where \qquad Ds = \log(N_r)/\log(1/r) \tag{16}$$

The ranges in the value of *Ds* characterize the type of fractal. A few methods to compute the fractal dimension have been published, such as Walking-Divider, Box Counting, Prism Counting, Epsilon-Blanket, Perimeter-Area Relationship, Fractional Brownian Motion, Power Spectrum, and Hybrid Methods [26]. For example, a simple wavelet-based fractal feature extraction algorithm has been used to recognize similar objects with very high accuracy [27].

Based on successful applications in medical image processing and recognition [28, 29], we proposed a novel feature extraction scheme: Firstly, the ROI of the calcification image is decomposed using 2D-wavelet transformation to create four decomposition subimages: $\boldsymbol{C_{LL}}$, $\boldsymbol{D_{LH}}$, $\boldsymbol{D_{HL}}$ and $\boldsymbol{D_{HH}}$. Then, the coefficients *I(x, y)* of each of the four wavelet subimages are used to calculate fractal number as follows:

$$d(k) = \sum_{x=0}^{M-1}\sum_{y=0}^{N-k-1} |I(x,y) - I(x,y+k)| / (M*(N-k)| + \sum_{y=0}^{N-1}\sum_{x=0}^{M-k-1} |I(x,y) - I(x+k,y)| / (N*(M-k) \tag{17}$$

and

$$F(k) = \log(d(k+1)) - \log(d(1)) \tag{18}$$

where, *k*=1,2,… *l*; *l* is number of feature. A vector of [*F (1), F (2), . ., F (l)*] will be fed into ANN classifier for training and testing.

IV. Classification

A three-layer ANN is used as a classifier. The layout of the ANN classifier is listed as follows:

No. of Input Layer: No. of features

No. of Hidden Layer: 50~100, depending on number of training samples

No. of Output Layer: 2 (Cancer and Normal Case)

Backpropagation algorithm is used in the ANN training procedure.

V. Experiment Results

In order to test the system's flexibility and MCCs detection performance, we conducted two experiments: one is to test classifier's convergence performance on the training set and to evaluate MCCs detection performance on the testing samples by inputting

different feature sets. Another experiment is to estimate MCCs' ROC curve on the testing samples.

One of databases is miniDDSM Database, which was created based on the original DDSM database [31] by the System Research Institute (SRI) at Alcorn State University, USA. The mammograms in the miniDDSM Database have been down-sampled in size. The image intensity is scaled into 256 grey levels.

A training set of 2,000 subimages with size of 128x128 are extracted from the subimages of microcalcification ROIs and the subimages of non-cancer areas. All of the subimages are manually selected from the mniDDSM database. Another set of 100 subimages of ROIs and 100 subimages of non-calcification areas are selected for the test.

In the first experiment, three sets of features (SRDM feature set, SRDM + Wavelet feature set, and SRDM + Wavelet-based fractal feature set) are used to train three ANN classifiers with the same configuration, respectively.

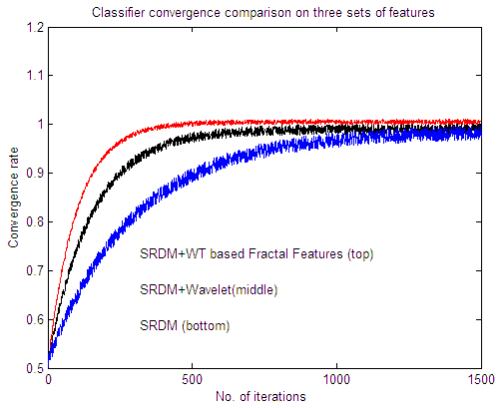

Fig. 4. ANN training convergence comparison on three sets of features

An ANN classifier trained by the third feature set (hybrid feature set: SRDM+ Wavelet-based fractal feature) can achieve the highest detection rate compared to the other two classifiers.

In our experiment, the wavelet feature set is extracted from 4x4 Daubechies wavelet coefficients, which is the result of multilevel 2D wavelet decompositions on ROI (128x128 subimage).

As shown in Fig. 4, the third set of SRDM+Wavelet-based fractal features has the best classifier convergence performance.

In the second experiment, the ROC curves are drawn.

TP (true positive): patient with disease (calcification) is correctly diagnosed.

FN (false negative): patient without disease (calcification) is diagnosed as diseased.

ANN confidence values given by the array: [0.65 0.70 0.75 0.80 0.85] were used for obtaining variation in order to draw the FROC curve. All of the experiments were conducted on a PC computer with a CPU processor of 2.5 GHz. The average speed for detecting a mammogram image is 5 seconds, including image preprocessing, feature extraction, and detection.

We tested on two databases: Mini-MIAS and miniDDSM. The FROC curves obtained are shown below:

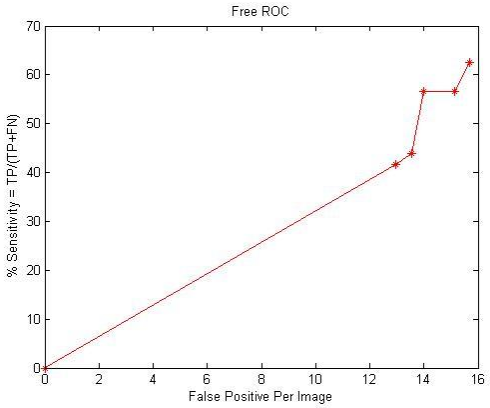

Fig. 5. Calcification detection performance on Mini-MIAS database (using SRDM + Wavelet-based fractal feature set)

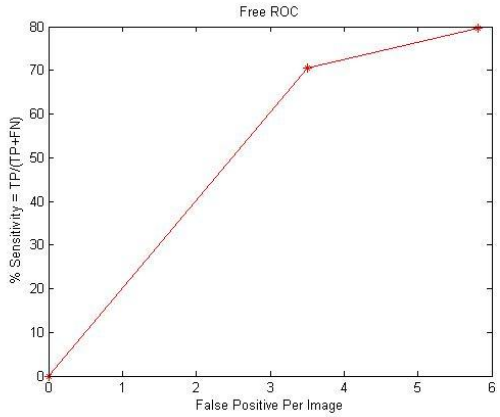

Fig. 6. Calcification detection performance on miniDDSM database (using SRDM + Wavelet-based fractal feature set)

VI. Conclusion

A novel hybrid feature extraction method is proposed and successfully applied to the detection of microcalcifications in digital mammograms. The hybrid feature set consists of the surrounding region dependence based features and wavelet-based fractal features. A merit of the proposed feature extraction method is that wavelet transformation can decompose

a ROI image into different subimages with different frequency bands and direction orientations, which means that the finer feature components may be explored in the 2D-WT. The cascaded fractal feature extraction based on the 2D-WT can overcome some deficiency of the 2D-WT.

The comparative experiments have demonstrated that the proposed feature extraction scheme has the best classifier training convergence performance among three sets of features used in the experiments and the FROC is also reported. Future work will focus on constructing a classification system with ensemble classifiers with hybrid features in order to increase system's reliability and detection rate at the same time in one system.

ACKNOWLEDGEMENTS

Part of this research is supported by Department of Defense (DOD TATRC), USA with the project of Development of a Knowledge Base to Support Detection and Diagnosis and Research in Mammography.

Two mammogram databases: miniDDSM database and Mini-MIAS database were used in the project. The miniDDSM was down-sampled from the Digital Database for Screening Mammography (DDSM) of University of South Florida [31]. Authors in this paper wish to thank professors and scientists who created and maintained the databases for this research. Authors sincerely thank Miss Vanessa Huston for proofreading and Dr. Collin D'Souza for his help in making FROC curves.

REFERENCES

[1] H. D. Cheng, X. Cai, X. Chen, L. Hu, X. Lou, Computer-aided detection and classification of microcalcifications in mammograms: A survey, Pattern Recognition, Vol. 36, No. 12, 2003, pp. 2967-2991.

[2] L. Wei, Y. Yang, R. M. Nishikawa, Y. Jiang, A study on several machine-learning methods for classification of malignant and benign clustered microcalcifications, IEEE Transactions on Medical Imaging, Vol. 24, No. 3, 2005, pp. 371-380.

[3] A. Papadopoulos, D. I. Fotiadis, A. Likas, An automatic microcalcification detection system based on a hybrid neural network classifier, Artificial Intelligence in Medicine, Vol. 25, 2002, pp.149-167.

[4]R. Panchal, B. Verma, Neural-association of Microcalcification patterns for their reliable classification in digital mammography, International Journal of Pattern Recognition and Artificial Intelligence, Vol. 20, No. 7, 2006, pp.971-983.

[5] I. El-Naqa, Y. Yang, M. N. Wernick, N. P. Galatsanos, R. M. Nishikawa, A support vector machine approach for detection of microcalcifications, IEEE Transactions on Medical Imaging, Vol. 21, No. 12, 2002, pp.1552-1563.

[6] J. Ye, S. Zheng, C. C. Yang, SVM-based microcalcification detection in digital mammograms, Proceedings of int. conf. on computer science and software engineering 6, 2008, pp.89-92.

[7] M. Haindl, S. Mikes, G. Scarpa, Unsupervised detection of mammogram regions of interest, Lecture notes in computer science, knowledge-based intelligent information and engineering systems, Springer Berlin/Heideberg, 2010, 4694, pp. 33-40.

[8] B. W. Hong, B. S. Sohn, Segmentation of regions of interest in mammogram in a topographic approach, IEEE Transactions on Information Technology in Biomedicine, Vol. 14, 2009, pp. 129-139.

[9] D. J. Marchette, R. A. Lorey, C. E. Priebe, An analysis of local feature extraction in digital mammography, Pattern Recognition, Vol. 30, No. 9, 1997, pp.1547-1554.

[10] L. Shen, R. M. Rangayyan, J. E. L. Desautels, Application of shape analysis to mammographic calcifications, IEEE Transactions on Medical Imaging, Vol. 13, No. 2, 1994, pp. 263-274.

[11] K. J. Kim, H. W. Park, Statistical textural features for detection of microcalcifications in digital mammograms. IEEE Transactions on Medical Imaging, Vol. 18, No. 3, 1999, pp. 231-238.

[12] S. N. Yu, K. Y. Li, Y. K. Huang, Detection of microcalcifications in digital mammograms using wavelet filter and Markov random field model, Computerized Medical Imaging and Graphics, Vol. 30, No. 3, 2006, pp. 163-173.

[13] R. N. Strickland, H. II. Hahn, Wavelet transforms for detecting microcalcification in mammograms, IEEE Transactions on Medical Imaging, Vol. 15, No. 2, 1995, pp. 218-229.

[14] T. C. Wang, N. B. Karayiannis, Detection of microcalcifications in digital mammograms using wavelets. IEEE Transactions on Medical Imaging, Vol. 17, No. 4, 1998, pp. 498-509.

[15] C. B. R. Ferreira, D. L. Borges, Analysis of mammogram classification using a wavelet transform decomposition, Pattern Recognition Letters, Vol. 24, No. 7, 2003, pp. 973-982.

[16] G. Boccignone,A. Chianese, A. Picariello, Computer aided detection of microcalcifications in digital mammograms, Computers in Biology and Medicine, Vol. 30, No. 5, 2000, pp. 267-286.

[17] S. Halkiotis, T. Botsis, M. Rangoussi, Automatic detection of clustered microcalcifications in digital mammograms using mathematical morphology and neural networks, Signal Processing, Vol. 87, No. 7, 2007, pp. 1559-1568.

[18] T. Stojic, I. Reljin, B. Reljin, Adaptation of multifractal analysis to segmentation of microcalcifications in digital mammograms, Physica A, Vol. 357, 2006, pp. 494-508.

[19] D. R. Chen, R. F. Chang, C. J. Chen, M. F. Ho, S. J. Kuo, S. T. Chen, S. J. Hung, W. K. Moon, Classification of breast ultrasound images using multifractal feature, Journal of Clinical Imaging, Vol. 29, 2005, pp. 235-245.

[20] L. Bocchi, G. Coppini, J. Nori, G. Vall, Detection of single and clustered microclfications in mammograms using multifractals models and neural networks, Medical Engineer & Physics, Vol. 26, 2004, pp. 303-312.
[21] M. D. Santo, M. Molinara, F. Tortorella, M. Vento, Automatic classification of clustered microcalcifications by a multiple expert system, Pattern Recognition, Vol. 36, No. 7, 2003, pp. 1467-1477.
[22] H. D. Cheng, J. L. Wang, X. J. Shi, Microcalcification detection using fuzzy logic and scale space approaches, Pattern Recognition, Vol. 37, No. 2, 2004, pp. 363-375.
[23] M. J. Bottema, J. P. Slavotinek, Detection and classification of lobular and DCIS (small cell) microcalcifications in digital mammograms, Pattern Recognition Letters, Vol. 21, No. 13-14, 2000, pp.1209-1214.
[24] J. Xu and J. S. Tang, Detection of clustered microcalcification using an improved textual approach for computer aided breast cancer diagnosis system, CSI Communications, January 2008, pp.17-19.
[25] C. K. Chui, Wavelets: A mathematical tool for signal analysis, SIAM, Society for Industrial and Applied Mathematics, Philadelphia, 1997.
[26] M. J. Turner, J. M. Blackledge, and P. R. Andrews, Fractal geometry in digital imaging, Academic Press, 1998.
[27] P. Zhang, T. D. Bui, and C. Y. Suen, Recognition of similar objects using 2-D wavelet-fractal feature extraction, Proceedings of 16th international conference on Pattern Recognition, 2002, Quebec, Canada.
[28] D. R. Chen, R. F. Chang, C. J. Chen, M.F. Ho, S. J. Kuo, C. T. Chen, S. J. Hung, Classification of breast ultrasound images using fractal feature, Journal of Clinical Imaging, Vol. 29, 2005, pp. 235-245.
[29] E. L. Chen, P.C. Chung, C. L. Chen, H. M.Tsai, and C. I. Chang, An automatic diagnostic system for CT liver image classification, IEEE Trans Biomed Eng. Vol. 45, No. 6, 1998, pp. 783–94.
[30] R. C. Gonzalez and R. E. Woods, Digital Image Processing, 3rd version, Pearson Prentice Hall, 2008.
[31] M. Heath, K. Bowyer, D. Kopans, R. Moore and W. P. Kegelmeyer, 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, 2001,ISBN 1-930524-00-5.