

Classification of Benign and Malignant Breast Masses Based on Shape and Texture Features in Sonography Images

Fahimeh Sadat Zakeri · Hamid Behnam ·
Nasrin Ahmadinejad

Received: 6 August 2010 / Accepted: 1 November 2010 / Published online: 17 November 2010
© Springer Science+Business Media, LLC 2010

Abstract The purpose of this research was evaluating novel shape and texture feature efficiency in classification of benign and malignant breast masses in sonography images. First, mass regions were extracted from the region of interest (ROI) sub-image by implementing a new hybrid segmentation approach based on level set algorithms. Then two left and right side areas of the masses are elicited. After that, six features (Eccentricity_feature, Solidity_feature, DeferenceArea_Hull_Rectangular, DeferenceArea_Mass_Rectangular, Cross-correlation-left and Cross-correlation-right) based on shape, texture and region characteristics of the masses were extracted for further classification. Finally a support vector machine (SVM) classifier was utilized to classify breast masses. The leave-one-case-out protocol was utilized on a database of eighty pathologically-proven breast sonographic images of patients (forty-seven benign cases and thirty-three malignant cases) to evaluate our method. The classification results showed an overall accuracy of 95.00%, sensitivity of 90.91%, specificity of 97.87%, positive predictive value of 96.77%, negative predictive value of 93.88%, and Matthew's correlation coefficient of 89.71%. The experimental results declare that our proposed method is actually a beneficial tool for the diagnosis of the breast cancer and can provide a second opinion for a physician's decision or can be used for the medicine training especially when coupled with other modalities.

Keywords Breast mass · Sonography · Shape feature · Texture feature · Classification

Introduction

One of every ten women in Europe and one of every eight women in the United States are affected by breast cancer [1] and in conformity with the American National Cancer Institute, every three minutes, one woman is concerned by a breast tumor and every thirteen minutes, one woman dies due to this disease. The cause of this disease has remained unknown yet, but early detection is so significant for successful control of this cancer with low costs, and considerable decrease in the death rate [2].

The used frequently practical methods for the breast cancer diagnosis are self-examination, sonography, mammography and biopsy. Biopsy is the most precise method for breast cancer detection, but it is an invasive and costly procedure for patients and society [3] with the very low positive rate between approximately 10% and 31% [4]. After biopsy, mammography is more accurate than other methods, but the ionized radiation of mammography can increase the health risk for the both patients and physicians. Currently, sonography is an important alternative for mammography. It is more comfortable, safer, and faster than mammography and it is more sensitive to detect abnormalities in dense breasts, for women younger than 35 years old [5].

Sonography is operator-dependent and the interpretation of sonography images requires an experienced and expert radiologist. Hence many studies have been done to introduce computer-aided diagnostic (CAD) systems in order to reduce unnecessary biopsies and assist inexperienced radiologists in diagnosis [5].

F. S. Zakeri · H. Behnam (✉)
Iran University of Science and Technology,
Tehran, Iran
e-mail: behnam@iust.ac.ir

N. Ahmadinejad
Tehran University of Medical Sciences,
Tehran, Iran

The previous researches in the field of CAD systems for the breast cancer in ultrasound images has been introduced and presented by Chang et al. [6, 7], Kuo et al. [8, 9], Chen et al. [5, 10–12], Horsch et al. [13], Mogatadakala et al. [14], Shankar et al. [15], and Behnam et al. [16].

In these researches, the mass region was located in the region of interest (ROI) area and features were extracted from this sub-image. In most cases, benign masses are well circumscribed and compact with round or ellipsoid shape and have smooth borders and homogeneous internal echoes and malignant masses have branch patterns, irregular shapes, spiculations, microlobulation, indistinct and angular borders with heterogeneous internal echoes [17].

The most applied texture features are auto-covariance coefficients [6, 11], occurrence matrix descriptors features [17, 18], posterior shadow or posterior echo [19], heterogeneous echo texture [20], and fractal dimension [18]. The global measures to describe morphological characteristics of the masses which are utilized in these systems are non-circumscribed or spiculated margins [20], normalized radial gradient [13], number of lobulations [21], and the depth to width ratio [21, 22].

One of the best results in this area was presented by Chang et al. [7]. They reported excellent classification results by applying the shape features (form factor, roundness, aspect ratio, convexity, solidity, and extent), that achieved overall sensitivity 88.89%, specificity 92.50%, positive predictive value 89.89% and negative predictive value of 91.74%.

In our study, six novel features (Eccentricity_feature, Solidity_feature, DeferenceArea_Hull_RectangularX, DeferenceArea_Mass_RectangularX, Cross-correlation-left and Cross-correlation-right) with an automatic segmentation are extracted and a support vector machine (SVM) classifier is utilized to categorize the breast masses to benign or malignant classes. The introduced features have never been employed in the breast mass classification previously, but in the micro calcification classification field, the first proposed feature was used with different meaning for the characterization of the shape of micro calcification clusters in [23].

In this study, we present novel shape features which can characterize breast sonographic masses better than the proposed features in other researches, and introduce two other regions which contain information about behavior of masses, specially their effects on surrounding tissues image. Extraction of previous texture features, like spatial gray level dependence matrices features, from these regions and their combination with other features also can be used in breast mass classification. These regions are considered by physicians in diagnosis, but haven't been used in CAD systems. Cross correlations of these regions are utilized because of smaller dependency on parameter adjustment of

sonographic machines. The obtained results declare that our proposed system is actually a beneficial tool for the diagnosis of the breast cancer and can be used as a secondary opinion for a physician's decision or for the medicine.

Subject and methods

Study subjects and image data acquisition

A breast pathologically -proven sonographic image database containing eighty cases (thirty-three malignant and forty-seven benign) was used for evaluating the performance of our method. The smallest diameter of each mass was larger than 1 cm. The age range of patients was between 18 to 64 years.

These images were recorded from February 19, 2006 to August 22, 2008 by using an Antares sonographic machine (Siemens Company, Germany) with a VFX 13–5 (Multi-D) linear array transducer by The Radiology Department of Imam Khomeini Hospital, Tehran, Iran.

Segmentation and mass extraction

Segmentation step is almost always necessary for feature extraction, and may affect strongly on the quality of the elicited features. Whereas some of the most important information is found in the transition region between the mass region and its background surrounding tissue, if the segmentation method cannot correctly identify the boundary between a mass and its background, the classification features may not comprise the significant transition region information [24].

Segmentation techniques face a great deal of challenges due to the characteristics of the ultrasound images: speckle noise, poor image contrast, weak boundaries and boundary gaps [17].

We introduce a new strategy which combines the gradient vector flow (GVF) and geodesic active contour (GAC) models which were developed by Paragios et al. [25] and Corsi et al. [26] respectively. The benefit of this method is that it permits for relatively free and simple initialization of the deformable model, while avoiding leakage at gap boundaries and weak edges. The formulation of this method is as follow:

$$\frac{\partial \varphi}{\partial t} = g \varepsilon \kappa |\nabla \varphi| - (1 - s) [\beta_1 (\vec{u}(x, y), \vec{v}(x, y)) \cdot \nabla \varphi + s \beta_2 \nabla g \nabla \varphi] \quad (1)$$

Where ε , β_1 and β_2 denote constant coefficients, φ is initial contour, t shows iteration, ∇ is the gradient operator

and g is a border stopping function that is approximately one in homogeneous regions and is close to zero on the object edges. κ is curvature function, the vector field $[\vec{u}, \vec{v}]$ is the GVF field, and (x, y) are the coordinate axes [27].

S function operates as a function of the edge map and iteration (time) named activation function calculated as follow:

$$s(f(x, y), t) = \begin{cases} 0 & \text{if for all } t' < t : \\ & E_{\varphi(x,y,t')=0}(f(x, y)) < T_{thresh} \\ 0.5 & \text{if for all } t' < t : \\ & T_{thresh} < E_{\varphi(x,y,t')=0}(f(x, y)) < \\ & T_{thresh} + 2 \\ 0.9 & \text{otherwise} \end{cases} \quad (2)$$

Where f is the edge map function with x and y coordinates, and E index denotes expectation value. T_{thresh} is the threshold value, and t shows iteration.

This function changes the effect of each part in the Eq. (1) with evolution. In the homogenous region, the average value of the edge map function is less than the threshold value therefore the second term in (1) equation which shows the external filed leads to converging the initial curve to the subject boundaries with the appropriate speed.

With closing the curve to the subject boundaries, in the evaluation process, sometime some parts of the curve are close to the object boundaries, while other parts is far from these boundaries. Now if edges that are close to evaluated curve are weak with gaps, high leakages can occur in these parts. The second and third terms of equation with S value 0.5 can help not to leak in weak

edges. Not passing across the weak edges is afforded by adding stronger stopping term which forces the evaluated curve into mass boundaries, the third term of the (1) equation.

Finally, in the vicinity of object boundaries, S value will be 0.9. Hence the effect of second part of Eq. (1) that shows external field is negligible. So the high power of the third term can evince an appropriate stop on the actual boundaries. This definition of activation function can present a satisfied segmentation with appropriate control in gaps and weak edges.

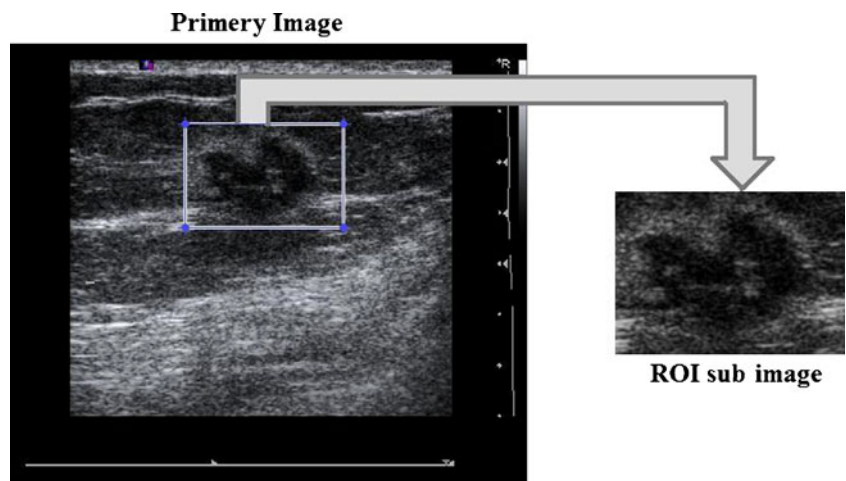
Value 0.9 in the activation function was chosen because of appropriate speed and control of leakage in evaluation process. The threshold value was calculated by the two-third of the edge map function average value at the first iteration by experience [28].

In this research, vector field convolution (VFC) is utilized instead of GVF field as external forces to evolve the initial contour. It has been demonstrated that the VFC snakes converge to boundary concavities the same as the GVF snakes. In addition, it needs fewer computations and is more robust to noise with better initialization [29].

In the segmentation step, the size of breast images are big, therefore the boundary curve evaluation can takes a long time. To get a satisfied speed in this step, before selecting an initial curve, the region of interest (ROI) sub-image that contains mass region and a little of the surrounding tissues were chosen by a radiologist then explained segmentation algorithm was applied to these sub-images.

An example of selecting a ROI sub-image is shown in Fig. 1, and two examples of the extracted ROI sub-images with results of the segmentation algorithm are displayed in Figs. 2 and 3 for a benign and malignant mass respectively in the breast sonographic images.

Fig. 1 ROI sub_image selected from primary image for a malignant mass



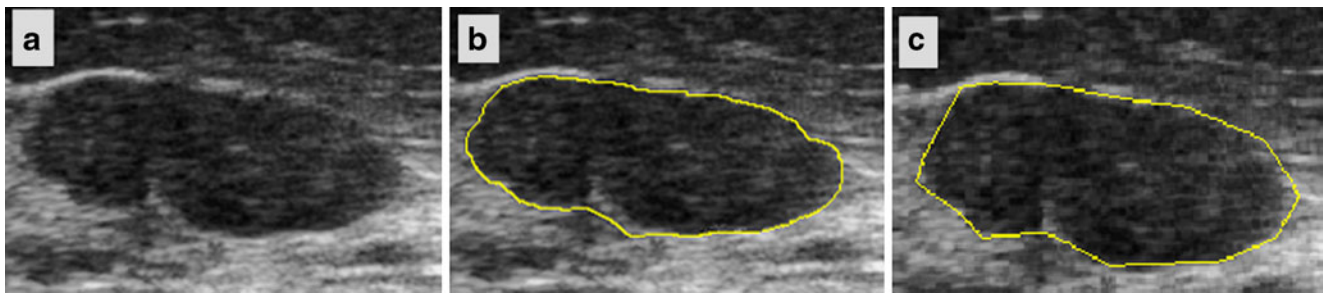


Fig. 2 a) An example of benign mass, b) automatic mass segmentation, c) manual mass segmentation by radiologist

Feature extraction

In general, feature extraction and selection are significant stages in supervised learning classification methods and particular features are required in any image processing and analysis applications.

The purpose of the feature extraction and selection is to maximize the distinguishing performance capability of them on the interested database [13], hence The considerations for choosing significant features principally include discrimination, optimality, reliability and independency [30]. For certain classification methods, like support vector machine and artificial neural networks, the dimension of feature space not only affects on their performance, but also determines the algorithm training time so that using too many features increase the classifier complexity and reduces the system performance and using not enough features increases miss-classification results.

In recent years, many kinds of features have been reported for breast mass classification like texture based features, region based features, image structure features, position related features, and shape-based features which are described and utilized in the CAD systems [31]. In our study a set of six novel features based on texture, region, and shape characteristics were extracted from the original images which are not applied in breast mass classification before and are defined as follows:

1. Eccentricity-feature: It is a scalar that describes the eccentricity of the ellipse which has the same second moment as the mass region.

The eccentricity value is obtained from the ratio of the distance between the foci of this ellipse to its major axis length.

2. Solidity-feature: It is a scalar that describes the proportion of the pixels in the mass region to the pixels in the convex hull including the mass. It is computed as (Mass's area)/ (Convex hull's area).
3. DifferenceArea-Hull-RectangularX: It is absolute value of convex hull's area minus convex RectangularX's area.
4. DifferenceArea-Mass-RectangularX: It is absolute value of convex RectangularX's area minus mass's area.
5. Cross-correlation-left: It is the cross correlation value between the RectangularX region and the Left-side region.
6. Cross-correlation-right: It is the cross correlation value between the RectangularX region and the Right-side region.

In Fig. 4 the convex hull which is the smallest convex polygon that contains the mass region is displayed. Mass's area is the actual number of pixels in the mass region and convex hull's area is the actual number of pixels in the convex hull region. RectangularX region is the smallest convex rectangular that its corner pixels are calculated from the coordinates of the most external pixels of the mass boundaries in y axis and x axis. According to the Fig. 5, RectangularX region's corner pixels of the segmented mass are calculated as below:

$$\begin{aligned} M &= (c_1, a_2), & N &= (d_1, a_2), & L &= (c_1, b_2), \\ P &= (d_1, b_2) \end{aligned} \quad (3)$$

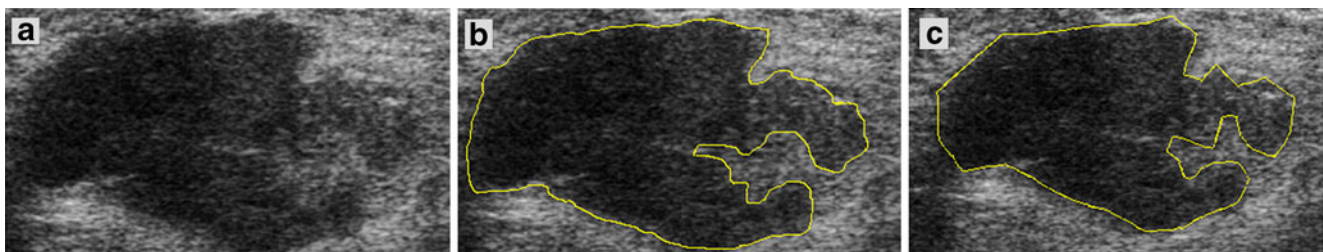


Fig. 3 a) An example of malignant mass, b) automatic mass segmentation, c) manual mass segmentation by radiologist

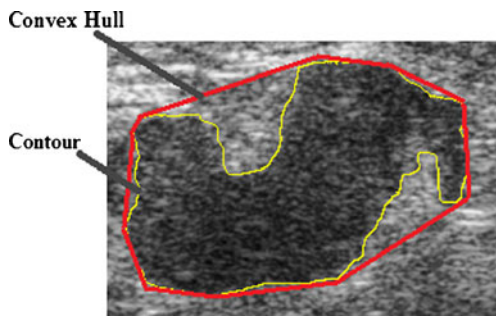


Fig. 4 An example of malignant mass with its Convex Hull- Yellow line is mass’s boundaries, Red line is Convex Hull’s boundaries

Left-side region and Right-side region mentioned in cross-correlation features are the regions that have the same depth and maximum half-width as the mass region in the left and right sides of the mass respectively. Figure 6 shows these regions.

Mass classification by support vector machine classifier

Support vector machine learning is a powerful machine learning tool which is capable of representing non-linear relationships of the features and to producing models that generalize well to a wide range of real-world applications including recognition, speaker identification, face detection, text categorization [23].

The basic idea of SVM classifier in pattern classification field can be expressed briefly as follows. First, the input vector is mapped into a feature space, either linearly or non-

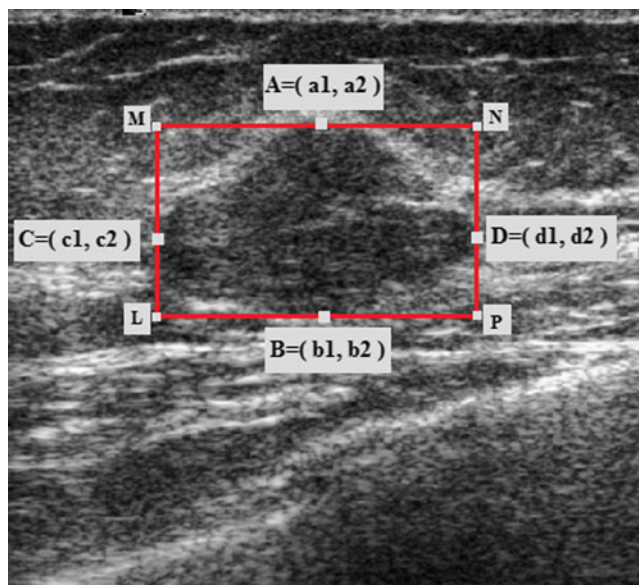


Fig. 5 An example of benign mass with its RectangularX region- Red line is RectangularX boundaries by (M, N, L, P) corner points; (A, B, C, D) are the most external points of the mass contour

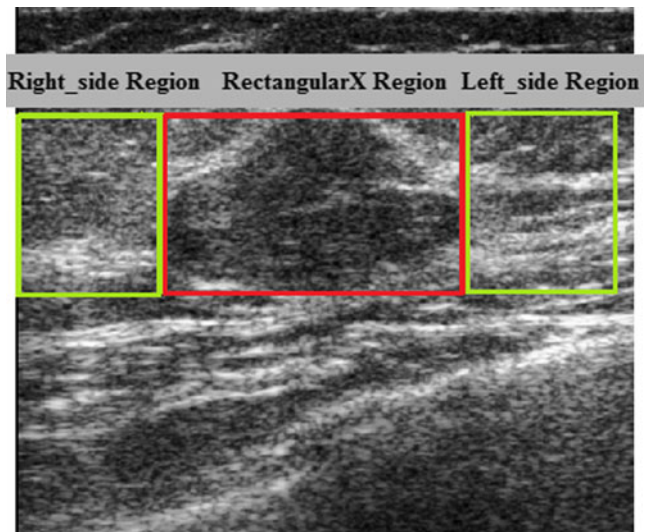


Fig. 6 An example of benign mass with Right_side, RectangularX, Left_side Regions

linearly, which is related to the kernel function selection. Then, the feature space is divided into two classes by the best constructed hyper plane by support vectors.

Lots of variety kernels can be employed for mapping function that has strong effect on the classification results, but the frequently used, useful kernel functions are Gaussian, Polynomial and Sigmoid kernels [31].

In our study, the Gaussian radial basis function (RBF) kernel was utilized because of its excellent results reported in many practical problems. Best results are obtained by sigma value -1.5 for leave-one-out protocol. The extracted features are applied as an input vector for this classifier in order to categorize the masses into benign and malignant cases. A complete description of SVMs theory for the pattern recognition field can be found in [32].

Result

In our experiments, we used a database of pathologically-proven digital breast sonographic images collected by the Department of Radiology at Imam Khomeini hospital. The analysis on this database were implemented in Matlab 2008 (MathWorks, Inc., Natick, MA), and results were achieved by testing the system with the leave-one-case-out protocol [33].

In the training phase of the classifier, in leave-one-case-out protocol, one case was leaved out in turn, while remaining cases were utilized to train the classifier; then the trained classifier were applied to classify the leave-out case. This procedure was repeated, until the classifier was tested on all cases in the dataset. Finally; these results were used for evaluating system’s performance.

Six objective indices—accuracy, sensitivity, specificity, positive predictive value, negative predictive value and Matthew's correlation coefficient [30]—were utilized to appraise the performance of our proposed system.

$$1. \text{ Accuracy} = (TP + TN)/(TP + TN + FP + FN).$$

$$2. \text{ Sensitivity} = TP/(TP + FN).$$

$$3. \text{ Specificity} = TN/(TN + FP).$$

$$4. \text{ Positive predictive value} = TP/(TP + FP).$$

$$5. \text{ Negative predictive value} = TN/(TN + FN).$$

$$6. \text{ Matthew's correlation coefficient} \\ = (TP \times TN - FP \times FN)/[(TP + FP)(TP + FN) \\ \times (TN + FP)(TN + FN)]^{0.5}$$

Where TP index shows the number of true positive cases, TN index shows the number of true negative cases, FP index indicates the number of false positive cases and FN index indicates the number of false negative cases.

The last formula, Matthew's correlation coefficient (MCC), has rarely been utilized for breast CAD system performance evaluation, but it is a powerful accuracy evaluation measure particularly when the numbers of negative and positive cases are unbalanced and it can give a better evaluation rather than overall accuracy [30].

Table 1 shows the performance of the SVM classifier for leave-one-out protocol. The percentage values of six evaluation indexes are shown in Table 2. We implemented the introduced features in [7] by Chang et al. on our database and the obtained results are also shown in Table 3.

The attained results declare the high discrimination ability of the proposed features. Hence, we can conclude

Table 1 Classification of breast tumors with the SVM classifier

Sonographic classification	SVM	
	Benign	Malignant
Benign	46 TN	3 FN
Malignant	1 FP	30TP
Total	47	33

Table 2 The objectively indices result for the proposed system

Index	SVM
Accuracy (%)	95.00
Sensitivity (%)	90.91
Specificity (%)	97.87
Positive predictive value (%)	96.77
Negative predictive value (%)	93.88
Matthew's correlation coefficient (%)	89.71

that our proposed approach is indeed a beneficial tool for diagnosis of breast cancer and medicine training.

Discussion

Many CAD systems have been recommended in breast mass classification field which utilized a wide range of features (morphological features, textural features and etc). In our study, a combination of the textural features which are less dependent on the transfer function of sonographic machines and machine's settings with shape features are introduced. In feature extraction the accuracy value of the segmentation algorithm is so important. We got approximately to the same results by automatic and manual segmentations that it shows the high accuracy of the proposed segmentation method.

In this study, the first four features based on shape characteristics describe irregularity, elongation, extension, and spread manners of the masses in surrounded tissues, hence, these features are adequate to classification of most benign masses from malignant masses because of the oval and regular shape of benign masses as compared with malignant ones. The last two features is the cross correlation of the three regions with the same depth, hence changing on sonographic machine settings (like gain) at each depth is similar for these regions. These features

Table 3 The objectively indices result for the proposed system by Chang et al. [7]

Index	SVM
Accuracy (%)	87.50
Sensitivity (%)	84.85
Specificity (%)	89.36
Positive predictive value (%)	84.85
Negative predictive value (%)	89.36
Matthew's correlation coefficient (%)	74.21

illustrate how masses influence on the image of healthy surrounded tissues.

By achieved results, it can be expressed that proposed features are excellent for breast mass classification and it can be easily executed on commercial sonographic machines.

In the future research, other novel features can be extracted and evaluated in order to reach better classification results. Also introducing an optimal Left-side and Right-side regions extraction method for each mass will be useful. As well, investigating effects of other classifiers on output results can be considered as subjects in order to achieve a more precise and reliable system for clinical examinations.

References

- Bothorel, S., Meunier, B. B., and Muller, S. A., Fuzzy logic based approach for semilogical analysis of microcalcification in mammographic images. *Intell. Syst.* 12:819–848, 1997.
- Junior, G. B., Paiva, A. C., Silva, A. C., and Muniz de Oliveira, A. C., Classification of breast tissues using Moran's index and Geary's coefficient as texture signatures and SVM. *Comput. Biol. Med.* 39:1063–1072, 2009.
- Joo, S., Yang, Y. S., Moon, W. K., et al., Computer-aided diagnosis of solid breast nodules: Use of an artificial neural network based on multiple sonographic features. *IEEE Trans. Med. Imaging* 23:1292–1300, 2004.
- Bassett, L. W., Liu, T. H., Giuliano, A. E., et al., The prevalence of carcinoma in palpable vs. impalpable, mammographically detected lesions. *AJR.* 157:21–24, 1991.
- Chen, D. R., Chang, R. F., Kuo, W. J., et al., Diagnosis of breast tumors with sonographic texture analysis using wavelet transform and neural networks. *Ultrasound Med. Biol.* 28:1301–1310, 2002.
- Chang, R. F., Wu, W. J., Moon, W. K., et al., Support vector machines for diagnosis of breast tumors on US images. *Acad. Radiol.* 10:189–197, 2003.
- Chang, R. F., Wu, W. J., Moon, W. K., et al., Automatic ultrasound segmentation and morphology based diagnosis of solid breast tumors. *Breast Cancer Res. Treat.* 89:179–185, 2005.
- Kuo, W. J., Chang, R. F., Moon, W. K., et al., Computer-aided diagnosis of breast tumors with different US systems. *Acad. Radiol.* 9:793–799, 2002.
- Kuo, W. J., Chang, R. F., Cheng, C. L., et al., Retrieval technique for the diagnosis of solid breast tumors on sonogram. *Ultrasound Med. Biol.* 28:903–909, 2002.
- Chen, C. M., Chou, Y. H., Han, K. C., et al., Breast lesions on sonograms: Computer-aided diagnosis with nearly setting-independent features and artificial neural networks. *Radiology* 226(2):504–514, 2003.
- Chen, D. R., Chang, R. F., and Huang, Y. L., Breast cancer diagnosis using self-organizing map for sonography. *Ultrasound Med. Biol.* 26:405–411, 2000.
- Chen, D. R., Chang, R. F., Huang, Y. L., et al., Texture analysis of breast tumors on sonograms. *Semin. Ultrasound CT MRI* 21:308–316, 2000.
- Horsch, K., Giger, M. L., Venta, L. A., and Vyborny, C. J., Computerized diagnosis of breast lesions on ultrasound. *Med. Phys.* 29:157–164, 2002.
- Mogatadakala, K., Donohue, K., Piccoli, C., and Forsberg, F., Detection of breast lesion regions in ultrasound images using wavelets and order statistics. *Med. Phys.* 33(4):840–849, 2006.
- Shankar, P., Piccoli, C., Reid, J., Forsberg, J., and Goldberg, B., Application of the compound probability density function for characterization of breast masses in ultrasound B scans. *Phys. Med. Biol.* 50(10):2241–2248, 2005.
- Behnam, H., Zakeri, F. S., and Ahmadinejad, N., Breast mass classification on sonographic images on the basis of shape analysis. *J. Med. Ultrason.* 37(4):181–186, 2010.
- Liua, B., Cheng, H. D., Huang, J., Tian, J., Tang, X., and Liu, J., Fully automatic and segmentation-robust classification of breast tumors based on local texture analysis of ultrasound images. *Pattern Recognit.* 43:280–298, 2010.
- Shi, X. Mass detection and classification in breast ultrasound images. Thesis of doctorate degree in computer science. Utah State University, Logan, Utah, 2006.
- Chen, S., Cheung, Y., Su, C., Chen, M., Hwang, T., and Hsueh, S., Analysis of sonographic features for the differentiation of benign and malignant breast tumors of different sizes. *Ultrasound Med. Biol.* 23(2):188–193, 2004.
- Tian, J. W., Sun, L. T., Guo, Y. H., Cheng, H. D., and Zhang, Y. T., Computerized-aid diagnosis of breast mass using ultrasound image. *Med. Phys.* 34:3158–3164, 2007.
- Segyeong, J., Yoon, S. Y., Woo, K. M., and Hee, C. K., Computer-aided diagnosis of solid breast nodules: Use of an artificial neural network based on multiple sonographic features. *IEEE Trans. Med. Imag.* 23(10):1292–1300, 2004.
- Cho, N., Moon, W., Cha, J., Kim, S., Han, B., Kim, E., Kim, M., Chung, S., Choi, H., and Im, J., Differentiating benign from malignant solid breast masses: Comparison of two-dimensional and three-dimensional US. *Radiology* 240(1):26–32, 2006.
- Wei, L., Yang, Y., and Nishikawa, R. M., Microcalcification classification assisted by content-based image retrieval for breast cancer diagnosis. *Pattern Recognit.* 42:1126–1132, 2009.
- Dominguez, A. R., and Nandi, A. K., Toward breast cancer diagnosis based on automated segmentation of masses in mammograms. *Pattern Recognit.* 42:1138–1148, 2009.
- Paragios, N., Mellina-Gottardo, O., and Ramesh, V., Gradient vector flow fast geometric active contours. *IEEE Trans. Pattern Anal. Mach. Intell.* 26:402–407, 2004.
- Corsi, C., Saracino, G., Sarti, A., and Lamberti, C., Left ventricular volume estimation for real-time three-dimensional echocardiography. *IEEE Trans. Med. Imag.* 21:1202–1208, 2002.
- Yu, H., A 3D multi view freehand ultrasound reconstruction system using volumetric registration and geometric level set segmentation. Thesis for Doctorate, University of New Mexico, December, 2006.
- Zakeri, F. S., Behnam, H., and Ahmadinejad, N., Breast mass diagnosis in sonographic images by using features based on mass contour and shape. 17th Iranian Conference on Electrical Engineering, Tehran, Iran, 2009.
- Li, B., and Acton, S. T., Active contour external force using vector field convolution for image segmentation. *IEEE Trans. Image Process.* 16:2096–2106, 2007.
- Cheng, H. D., Shan, J., Ju, W., Guo, Y., and Zhang, L., Automated breast cancer detection and classification using ultrasound images: A survey. *Pattern Recognit.* 43:299–317, 2010.
- Subashini, T. S., Ramalingam, V., and Palanivel, S., Automated assessment of breast tissue density in digital mammograms. *Comput. Vis. Image Underst.* 114:33–43, 2010.
- Vapnik, V., *Statistical learning theory*. Wiley: New York, 1998.
- Hastie, T., Tibshirani, R., and Friedman, J., *The elements of statistical learning. Data mining, inference, and prediction*. Springer-Verlag: Berlin, 2001.