Journal of Computer Science 9 (6): 726-732, 2013
ISSN: 1549-3636

doi:10.3844/jcssp.2013.726.732 Published Online 9 (6) 2013 (http://www.thescipub.com/jcs.toc)

Mass Classification in Digital Mammograms based on Discrete Shearlet Transform

[1]Amjath Ali, J. and [2]J. Janet

[1]Department of Computer Science, St.Peter's University, Chennai, India
[2]Department of Computer Science, Dr. M.G.R. University, Chennai, India

Received 2013-05-03, Revised 2013-05-29; Accepted 2013-06-01

ABSTRACT

The most significant health problem in the world is breast cancer and early detection is the key to predict it. Mammography is the most reliable method to diagnose breast cancer at the earliest. The classification of the two most findings in the digital mammograms, micro calcifications and mass are valuable for early detection. Since, the appearance of the masses are similar to the surrounding parenchyma, the classification is not an easy task. In this study, an efficient approach to classify masses in the Mammography Image Analysis Society (MIAS) database mammogram images is presented. The key features used for the classification is the energies of shearlet decomposed image. These features are fed into SVM classifier to classify mass/non mass images and also benign/malignant. The results demonstrate that the proposed shearlet energy features outperforms the wavelet energy features in terms of accuracy.

Keywords: Wavelet Transform, Shearlet Transform, SVM Classifier, Digital Mammograms, Mass Region of Interest

1. INTRODUCTION

The most dangerous type of cancer is breast cancer. Over 11% of woman gets affected by breast cancer in their life time. Hence, an early detection of breast cancer is essential to reduce mortality. More studies are attempted to classify masses in mammograms in the last decade. Ensemble neural network along with K-Nearest Neighbor (KNN) classifier is presented in (McLeod and Verma, 2012) for classifying the masses in digital mammograms. This technique is mainly considered for improving the classification rate of multilayer perceptron network. Benchmark database is used for classifying the masses in digital mammogram.

Mass classification problems are investigated in (Hussain *et al*., 2012) with different Gabor features. Gabor feature extraction technique is used to extract multi-scale and multi-orientation texture features. This represents the structural properties of both mass and normal tissue in mammogram. For classifying images along with Gabor features successive enhancement learning based weighted support sector machine is used. Masses in mammogram are classified using Extreme Learning Machine (ELM) based classifier is proposed in (Vani *et al*., 2010). The database used for this classification is mini MIAS database. ELM algorithm mainly depends on number of hidden neurons and the magnitude of input weight initialization. In order to achieve the greatest classification performance the best numbers of neurons are identified.

Automatic mass classification into benign and malignant is presented in (Jasmine *et al*., 2011) based on features extracted from the contourlet coefficients using non subsampled contourlet transform, the mammogram images are classified by using a classifier based on support vector machine. Discrete wavelet transform decomposition and an artificial neural network classifier is used for malignant/benign masses classification of

Corresponding Author: Amjath Ali J., Department of Computer Science, St.Peter's University, Chennai, India

region of interest in mammograms is proposed in (Fraschini, 2011).

A novel opposition based classifier is developed in (Saki *et al*., 2010) which classify breast masses into benign and malignant categories. The classifier used is multi layer perceptron network with a novel learning rule. The features include circularity, Zernike moments, contrast and average gray level. Decision making is performed in both feature extraction and classification stage. The whole block presented in (Mencattini *et al*., 2010) represents the final step of a computer aided diagnostic system developed at the University of Rome 'Tor Vergata' which is able to assist the radiologist in the early diagnosis of breast cancer that includes masses segmentation, features extraction and selection and masses classification.

Images are separated into two groups for classifying mass images in (Martins *et al*., 2009), based on shape and texture descriptor i.e., masses and non masses. Shape descriptor includes eccentricity, circularity and convexity. For extracting the descriptors 4x4 window size is used to allow the calculation of co-occurrence. Artificial neural network is used to categorize the masses, which performs benign-malignant classification on region of interest that contains mass is explained in (Islam *et al*., 2010). Texture is the major characteristics of mass classification. The textural features used for characterizing the masses are mean, standard deviation, entropy, skewness, kurtosis and uniformity. Support vector machine and wavelet decomposition is investigated for classifying the masses in (Gorgel *et al*., 2009). Decision making is performed in both feature extraction and classification stage.

1.1. Methodology

The proposed system for the classification of mass in digital mammogram is developed based on DST and SVM classifier. The theoretical background of all the approaches are introduced.

1.2. Discrete Shearlet Transform

The proposed approach for the classification of microcalcification is based on a new multi resolution transform called the shearlet transform introduced by the authors in (Easley *et al*., 2008). An N*N image consists of a finite sequence of values, $\{x[n_1,n_2]_{n_1,n_2=0}^{N-1,N-1}\}$ where $N\in\mathbb{N}$. Identifying the domain with the finite group $\mathbb{Z}_N^2$, the inner product of image $x,y:\mathbb{Z}_N^2\to\mathbb{C}$ is defined as Equation 1:

$$(x,y)=\sum_{u=0}^{N-1}\sum_{v=0}^{N-1}x(u,v)\overline{y(u,v)} \tag{1}$$

Thus the discrete analog of $L^2(\mathbb{R}^2)$ is $l^2(\mathbb{Z}_N^2)$. Given an image $f\in l^2\left(\mathbb{Z}_N^2\right)$, let $\hat{f}[k_1,k_2]$ denote its 2D Discrete Fourier Transform (DFT) Equation 2:

$$\hat{f}[k_1,k_2]=\frac{1}{N}\sum_{n_1,n_2=0}^{N-1}f[n_1,n_2]e^{-2\Pi i(\frac{n_1}{N}k_1+\frac{n_1}{N}k_2)} \tag{2}$$

The brackets in the equations $[\cdot,\cdot]$ denote arrays of indices and parentheses $(\cdot,\cdot)$ denote function evaluations. Then the interpretation of the numbers $\hat{f}[k_1,k_2]$ as samples $\hat{f}[k_1,k_2]=\hat{f}(k_1,k_2)$ is given by the following equation from the trigonometric polynomial Equation 3:

$$\hat{f}(\xi_1,\xi_2)=\sum_{n_1,n_2=0}^{N-1}f[n_1,n_2]e^{-2\Pi i(\frac{n_1}{N}\xi_1+\frac{n_1}{N}\xi_2)} \tag{3}$$

First, to compute Equation 4:

$$\hat{f}(\xi_1,\xi_2)\overline{V(2^{-2j}\xi_1,2^{-2j}\xi_2)} \tag{4}$$

In the discrete domain, at the resolution level J, the Laplacian pyramid algorithm is implemented in the time domain. This will accomplish the multi scale partition by decomposing $f_a^{j-1}[n_1,n_2],0\le n_1,n_2<N_j-1$, into a low pass filtered image $f_a^j[n_1,n_2]$, a quarter of the size of $f_a^{j-1}[n_1,n_2]$ and a high pass filtered image $f_d^{j-1}[n_1,n_2]$. Observe that the matrix $f_a^{j-1}[n_1,n_2]$ has size N_j*N_j, where $N_j = 2^{-2j}N$ and $f_a^0[n_1,n_2]=f[n_1,n_2]$ has size N*N in particular Equation 5:

$$\hat{f}_d^j(\xi_1,\xi_2)=\hat{f}(\xi_1,\xi_2)\overline{V(2^{-2j}\xi_1,2^{-2j}\xi_2)} \tag{5}$$

Thus, $f_d^j[n_1,n_2]$ are the discrete samples of a function $f_d^j[x_1,x_2]$, whose Fourier transform is $\hat{f}_d^j(\xi_1,\xi_2)$. In order to obtain the directional localization the DFT on the

pseudo-polar grid is computed and then one-dimensional band-pass filter is applied to the components of the signal with respect to this grid. More precisely, the definition of the pseudo-polar co ordinates $(u,v)\in\mathbb{R}^2$ as follows Equation 6 and 7:

$$(u,v)=(\xi_1,\frac{\xi_2}{\xi_1}),\text{if}(\xi_1,\xi_2)\in D_0 \quad (6)$$

$$(u,v)=(\xi_1,\frac{\xi_1}{\xi_2}),\text{if}(\xi_1,\xi_2)\in D_1 \quad (7)$$

After performing this change of co ordinates, $g_j(u,v)=\hat{f}_d^j(\xi_1,\xi_2)$ is obtained and for l = 1-2^j,…^{2}-1 Equation 8:

$$\begin{aligned}\hat{f}(\xi_1,\xi_2)&=\overline{V(2^{-2j}\xi_1,2^{-2j}\xi_2)W_{jl}^{(d)}(\xi_1,\xi_2)}\\&=g_j(u,v)\overline{W(2^jv-l)}\end{aligned} \quad (8)$$

This expression shows that the different directional components are obtained by simply translating the window function w. The discrete samples $g_j[n_1,n_2]=g_j(n_1,n_2)$ are the values of the DFT of $f_d^j[n_1,n_2]$ on a pseudo-polar grid. That is, the samples in the frequency domain are taken not on a Cartesian grid, but along lines across the origin at various slopes. This has been recently referred to as the pseudo-polar grid. One may obtain the discrete Frequency values of f_d^j on the pseudo-polar grid by direct extraction using the Fast Fourier Trans-form (FFT) with complexity ON^2 logN or by using the Pseudo-polar DFT (PDFT).

1.3. SVM Classifier

In medical Support Vector Machines (SVMs) are a set of related supervised learning methods that analyze data and recognize patterns, used for classification and regression analysis (Rejani and Selvi, 2009). The standard SVM is a non-probabilistic binary linear classifier, i.e., it predicts, for each given input, which of two possible classes the input is a member of. A classification task usually involves with training and testing data which consists of some data instances. Each instance in the training set contains one "target value" (class labels) and several "attributes" (features) (Gorgel *et al*., 2009). SVM has an extra advantage of automatic model selection in the sense that both the optimal number and locations of the basic functions is automatic ally obtained during training. The performance of SVM largely depends on the kernel.

SVM is essentially a linear learning machine. For the input training sample set Equation 9:

$$(x_i,y_i),i=1....n,x\in R^n,y\in\{-1,+1\} \quad (9)$$

The classification hyperplane equation is Equation 10:

$$(\omega.x)+b=0 \quad (10)$$

Thus the classification margin is 2/|ω|. To maximize the margin, that is to minimize |ω|, the optimal hyperplane problem is transformed to quadratic programming problem as follows Equation 11:

$$\begin{aligned}&\min\Phi(\omega)=\frac{1}{2}(\omega,\omega)\\&\text{s.t.}y_i((\omega.x)+b\geq 1,i=1,2....l\end{aligned} \quad (11)$$

After introduction of Lagrange multiplier, the dual problem is given by Equation 12:

$$\begin{aligned}&\max Q(\alpha)=\sum_{i=1}\alpha_i-\frac{1}{2}\sum_{i=1}\sum_{i=1}y_iy_j\alpha_i\alpha_jK(x_ix_j)\\&\text{s.t}\sum_{i=1}^{n}y_i\alpha_i=0,\alpha_i\geq 0,i=1,2,....,n\end{aligned} \quad (12)$$

According to Kuhn-Tucker rules, the optimal solution must satisfy Equation 13:

$$\alpha_i(y_i((w.x_i)+b)-1=0,i=1,2,...n \quad (13)$$

That is to say if the option solution is Equation 14:

$$\alpha^*=(\alpha_1^*,\alpha_2^*,....,\alpha_i^*)^T,i=1,2,...n \quad (14)$$

Then Equation 15:

$$\begin{aligned}&w^*=\sum_{i=1}^{n}\alpha_i^*y_ix_i\\&b^*=y_i-\sum_{i=1}^{n}y_i\alpha_i^*(x_i.x_j),j\in\{j\,|\,\alpha_i^*\rangle 0\}\end{aligned} \quad (15)$$

For every training sample point x_i, there is a corresponding Lagrange multiplier and the sample

points that are corresponding to $\alpha_i = 0$ don't contribute to solve the classification hyperplane while the other points that are corresponding to $\alpha_i > 0$ do, so it is called support vectors. Hence the optimal hyperplane equation is given by Equation 16:

$$\sum_{x_i \in SV} \alpha_i y_i (x_i . x_j) + b = 0 \quad (16)$$

The hard classifier is then Equation 17:

$$y = \text{sgn}\left[\sum_{x_i \in SV} \alpha_i y_i (x_i . x_j) + b \right] \quad (17)$$

For nonlinear situation, SVM constructs an optimal separating hyperplane in the high dimensional space by introducing kernel function $K(x.y) = \phi(x).\phi(y)$, hence the nonlinear SVM is given by Equation 18:

$$\begin{aligned} &\min \Phi(\omega) = \frac{1}{2}(\omega, \omega) \\ &\text{s.t.} y_i((\omega.\varphi(x_i)) + b) \geq 1, i = 1, 2, ...l \end{aligned} \quad (18)$$

And its dual problem is given by Equation 19:

$$\begin{aligned} &\max L(\alpha) = \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i=1}^{l} \sum_{i=1}^{l} y_i y_j \alpha_i \alpha_j K(x_i . x_j) \\ &\text{s.t.} \sum_{i=1}^{n} y_i \alpha_i = 0, 0 \leq \alpha_i \leq C. i = 1, 2, ..., l \end{aligned} \quad (19)$$

Thus the optimal hyper plane equation is determined by the solution to the optimal problem. The SVM classifier implementation is standard implementation.

In the MATLAB environment svmtrain and svmclassify functions are used to train and test the mammogram images.

1.4. Experimental Setup

The main objective of this study is to distinguish between the types of mass tumors as benign or malignant. MIAS database, a benchmark database used by many researchers is taken to evaluate this study. The original mammograms in MIAS are very big size (1024×1024 pixels). However, the whole image consists of noise and background around 50%. In order to remove this unwanted background and noise, a cropping operation is done manually before feature extraction by choosing the given center of abnormality as the center of the ROI. The size of the cropped ROI is 256×256. **Figure 1** shows the original mammogram image (mdb134) and **Fig. 2b** shows the cropped image.

The next step after cropping is feature extraction from the set of training cropped ROI's images. The shearlet transform is used to represent the cropped ROI's in multi-scale and multi-directions. The decomposition level from 2 to 4 with various directions from 2 to 64 is used in this study and the cropped images are transformed into the aforementioned levels and directions. Then the energy of each directional sub-band is extracted and used as feature vector for the corresponding training image.

Figure 2 shows the proposed mass classification system. The energy of each directional sub-band of the image I is calculated by using the formula in Equation 20:

$$\text{ENERGY}_e = \frac{1}{RC} \sum_{i=1}^{R} \sum_{j=1}^{C} \left| I_e(i, j) \right| \quad (20)$$

where, $I_e(I,j)$ the pixel value of the eth sub-band and R, C is width and height of the sub-band respectively.

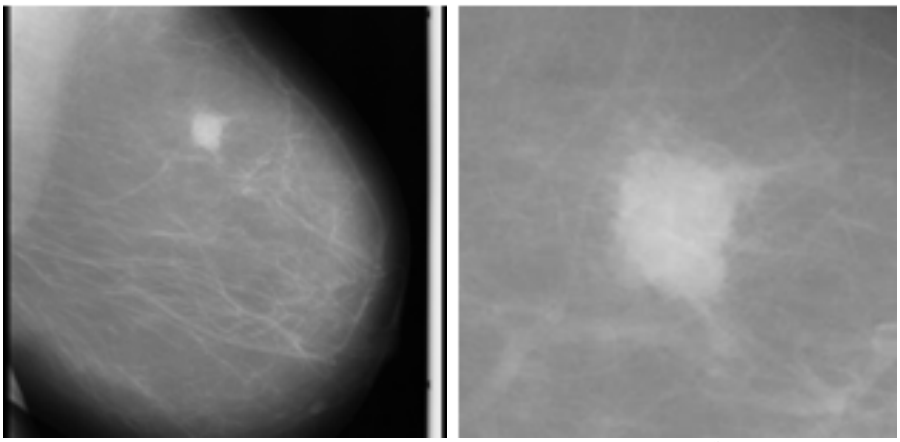
The feature vector is calculated for all training ROI's and stored in the feature database. This database is used to train the classifier. The final step is the classification stage.

The SVM classifier is used to classify the mass images. The kernel used in SVM is linear kernel. Initially, the SVM classifier is trained by the feature database. For an unknown mammogram image to be classified, the proposed shearlet energy features are extracted as don e in training stage and then the trained classifier is used to classify the severity of the given mammogram image.

As, the classification is done in two stages, two SVM classifier are used in this study. Both the classifiers are trained by using separate databases. The first SVM classifier classifies the given unknown mammogram image into normal or abnormal. If it is found abnormal, then the second SVM classifier classifies again the mass severity as benign/malignant.

1.5. Experimental Results

The evaluation is carried out in relation with two stages of mass classification system namely (i) normal/abnormal classification (ii) benign/malignant classification.



(a) (b)

Fig. 1. (a) original image (b) cropped image

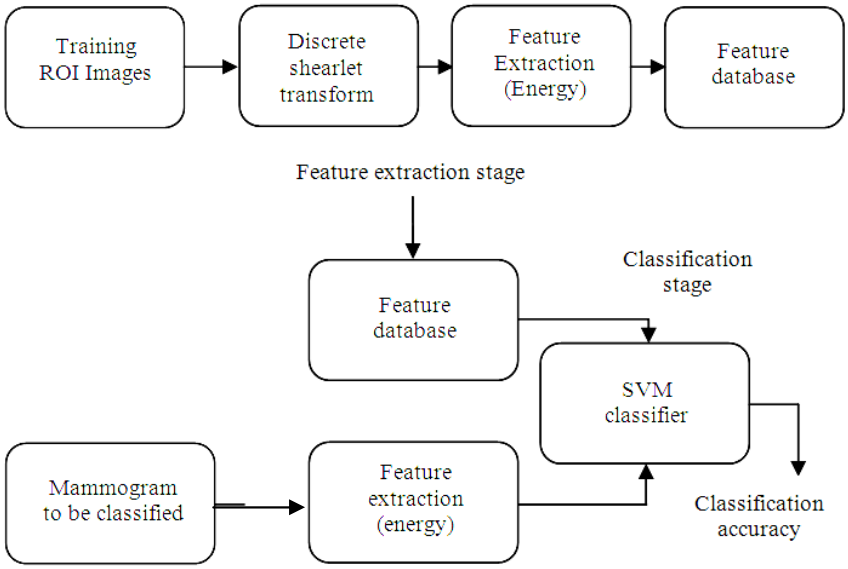


Fig. 2. Block diagram of the proposed approach for mass classification

This method is applied to a set of 126 (70 normal/56 abnormal) mass images of MIAS database. The images are in 8bit gray scale images and of size 1024×1024 pixels. The input to the proposed system is the suspicious area inside the whole mammogram MLO images. The definition of suspicious area that is the region of interest is given by the MIAS database.

First, the ROI which contain the suspicious area in the mammogram is extracted. In the second stage, the ROI is projected to the multi directional domain, shearlet transform and the designated energy features are extracted. Finally, linear SVM classifier is applied to classify the mammogram mass images. The numbers of mammogram images to train and test the SVM classifier are shown in **Table 1**. **Table 2** shows the classification accuracy for normal and abnormal while the classification accuracy for mass severity is shown in **Table 3**.

Table 1. Number of images used to train SVM Classifier

Images	Normal	Abnormal	Benign	Malignant
Training	47	37	25	13
Testing	70	56	37	19

Table 2. Classification accuracy of normal/abnormal cases based on DST

	Classification rate (%)					
	Level 2		Level 3		Level 4	
Number of directions	Normal	Abnormal	Normal	Abnormal	Normal	Abnormal
2	78.57	55.36	84.29	50.00	78.57	57.14
4	77.14	60.71	78.57	60.71	78.57	67.86
8	78.57	64.29	84.29	67.86	88.57	69.64
16	82.86	64.29	85.71	71.43	92.86	73.21
32	90.00	73.21	92.86	76.79	87.14	89.29
64	95.29	82.14	91.43	87.50	88.57	94.64

Table 3. Classification accuracy of benign/malignant cases based on DST

	Classification rate (%)					
	Level 2		Level 3		Level 4	
Number of directions	Benign	Malignant	Benign	Malignant	Benign	Malignant
2	86.49	31.58	86.49	36.84	86.49	52.63
4	94.59	52.63	100.00	52.63	91.90	63.16
8	94.59	57.89	94.59	68.42	91.90	73.68
16	94.59	73.68	100.00	68.42	100.00	78.95
32	97.30	84.21	97.30	89.47	97.30	89.47
64	100.00	89.47	100.00	89.47	94.60	94.74

Table 4. Classification accuracy of normal/abnormal cases based on DWT

	Classification rate (%)					
	Bior3.7		db8		Sym8	
Level	Normal	Abnormal	Normal	Abnormal	Normal	Abnormal
2	87.14	37.14	91.43	38.57	85.71	44.29
3	91.43	40.00	78.57	52.86	87.14	47.14
4	85.71	45.71	88.57	50.00	94.29	45.71
5	77.14	58.57	82.86	55.71	81.43	60.00

Table 5. Classification accuracy of benign/malignant cases based on DWT

	Classification rate (%)					
	Level 2		Level 3		Level 4	
Level	Benign	Malignant	Benign	Malignant	Benign	Malignant
2	91.89	36.84	100.00	31.58	97.30	42.11
3	94.59	42.11	94.59	36.84	91.89	42.11
4	91.89	52.63	83.78	57.89	10.00	31.58
5	91.89	52.63	81.08	73.68	83.78	57.89

As the number of directional sub-bands in DST depends on the size of the directional filter, the performance is evaluated by varying the direction filter size used in the decomposition stage. In order to analyze the efficiency of DST over wavelet transform, three wavelets bi-orthogonal (bior 3.7), Daubechies-8(db8) and Symlet (sym8) are used. As wavelet transform is a multi scale analysis, the level of decomposition is varied to obtain higher classification rate for both normal/abnormal and benign/malignant classification stages and the obtained accuracy is tabulated in **Table 4 and 5** respectively.

2. CONCLUSION

In this study, an efficient approach to classify digital mammogram images which contains masses based on DST and SVM classifier is presented. The energies extracted from the shearlet decomposed mammogram images are used as key features. To classify the mammogram images, these extracted features are fed to the SVM classifier as an input. The output from the SVM classifier classifies the given mammogram images into normal/abnormal initially and then the cancerous mammograms are again classified as benign/malignant. The performance metric used to evaluate the proposed approach is the classification accuracy. The results demonstrate that shearlet energy features based approach yielded better classification accuracy compared to wavelet energy features based approach.

3. REFERENCES

Easley, G., D. Labate and W.Q. Lim, 2008. Sparse directional image representations using the discrete shearlet transform. App. Comput. Harmon. Anal., 25: 25-46. DOI: 10.1016/j.acha.2007.09.003

Fraschini, M., 2011. Mammographic masses classification: Novel and simple signal analysis method. Proceedings of the Electronics Letters, Jan. 6, IEEE Xplore Press, pp: 14-15. DOI: 10.1049/el.2010.2712

Gorgel, P., A. Sertbas, N. Kilic, O.N. Ucan and O. Osman, 2009. Mammographic mass classification using wavelet based support vector machine. J. Elect. Elect. Eng., 9: 867-875.

Hussain, M., S. Khan, G. Muhammad and G. Bebis, 2012. A Comparison of different gabor features for mass classification in mammography. Proceedings of the 8th International Conference on Signal Image Technology and Internet Based Systems, Nov. 25-29, IEEE Xplore Press, Naples, pp: 142-148. DOI: 10.1109/SITIS.2012.31

Islam, M.J., M. Ahmad and M.A. Sid-Ahmed, 2010. An efficient automatic mass classification method in digitized mammograms using artificial neural network. Int. J. Artif. Intell. Applic., 1: 6-6. DOI: 10.5121/ijaia.2010.130

Jasmine, J.S.L., S. Baskaran and A. Govardhan, 2011. An automated Mass classification system in digital mammograms using contourlet transform and support vector machine. Int. J. Comp. Appl., 31: 54-54.

Martins, L.D.O., G.B. Junior, A.C. Silva, A.C.D. Paiva and M. Gattass, 2009. Detection of masses in digital mammograms using k-means and support vector machine. Elect. Lett. Comp. Vis. Image Anal.

McLeod, P. and B. Verma, 2012. Clustered ensemble neural network for breast mass classification in digital mammography. Proceedings of the International Joint Conference on Neural Networks, Jun. 10-15, IEEE Xplore Press, Brisbane, QLD, pp: 1-6. DOI: 10.1109/IJCNN.2012.6252539

Mencattini, A., M. Salmeri, G. Rabottino and S. Salicone, 2010. Metrological characterization of a cadx system forthe classification of breast masses in mammograms. IEEE Trans. Instrument. Measur., 59: 2792-2799. DOI: 10.1109/TIM.2010.2060751

Rejani, Y.A. and S.T. Selvi, 2009. Early detection of breast cancer using SVM classifier technique. Int. J. Comput. Sci. Eng., 1: 127-130.

Saki, F., A. Tahmasbi and S.B. Shokouhi, 2010. A novel opposition based classifier for mass diagnosis in mammography images. Proceedings of the 17th Iranian Conference of Biomedical Engineering, Nov. 3-4, IEEE Xplore Press, Isfahan, pp: 1-4. DOI: 10.1109/ICBME.2010.5704940

Vani, G., R. Savitha and N. Sundararajan, 2010. Classification of abnormalities in digitized mammograms using extreme learning machine. Proceedings of the 11th International Conference on Control Automation Robotics and Vision, Dec. 7-10, IEEE Xplore Press, Singapore, pp: 2114-2117. DOI: 10.1109/ICARCV.2010.5707794