Journal of ecology of Health & Environment
An International Journal

http://dx.doi.org/10.12785/jehe/030303

Computer Aided System for Breast Cancer Diagnosis in Ultrasound Images

Noura M. A. Abdelwahed, Mohamed Meselhy Eltoukhy * *and M. E. Wahed.*

Faculty of Computers and Informatics, Suez Canal University, Ismailia, 41522, Egypt.

Received: 17 May 2015, Revised: 17 Jul. 2015, Accepted: 20 Jul. 2015.
Published online: 1 Sep. 2015.

Abstract: Breast cancer is the top cancer in women both in the developed and the developing world. Unfortunately, most of breast cancer is diagnosed in very late stages. Present there is no known method to prevent breast cancer but early detection increase the chance of cure. Developing computer aided diagnosis system (CAD) can help radiologist in their decision. In this paper, we proposed a CAD system to segment and classify the breast cancer in ultrasound images. The system is using marker controlled watershed transformation technique to identify the region of interest (ROI). Then, wavelet transform is applied to extract set of features combined with texture and statistical features. The classification step determines whether the ROI is normal or focal lesion. Finally, focal lesion is classified as benign or malignant. Support vector machine (SVM), K-nearest neighbour (KNN) and classification&regression trees (CART) are used to achieve the classification task. The proposed method is validated using 10 folds cross validation and the obtained results were encouraging. The results show that CART obtain 83.75% classification rate using statistical and texture features in case of classify benign and malignant tumor which is more than SVM and KNN classification rate. In case of differentiate between normal and abnormal classes SVM and CART obtain 100% classification rate using texture feature.

Keywords: Breast cancer, Image segmentation, Wavelet transform, Feature extraction, Breast ultrasound image.

1 Introduction

Breast cancer is the disease that threatens women lives. Early detection prevents cancer spreading to all parts of the body [1] [2] [3]. Mammography plays an important role in this field [4]. Its limitations are the dependence on the composition of mammary parenchyma and tumor tissue [2] [5]. Ultrasound breast images use in addition to mammography to reduce the number of unnecessary biopsies for patients and improve the diagnostic accuracy rate [2] [5] [6] [7] [8]. Unnecessary biopsies put patient under pressure and worry. Computer Aided Diagnosis (CAD) systems are developed to help the radiologists in detecting the abnormal regions present in the breast. The CAD systems act as second reader, while the final decision lies with the radiologist.

Eltoukhy et al. [9] built their system on multiresolution representation of the mammogram images using curvelet and wavelet transform. They cropped the region of interest (ROI) manually. They used curvelet and wavelet transformation methods separately. The features are extracted from the cropped images based on a multiresolution transform. They applied four different decomposition level based on three different wavelet functions. The biggest 100 coefficients are extracted in order to represent the corresponding mammogram. These coefficients are sorted in descending order. They used nearest neighbor based on Euclidian distance in classification step. The extracted features based on curvelet give a better performance as compared to wavelet. For classification of normal, benign and malignant based on curvelet transform coefficients; the classifier achieved 94.07%, while the highest rate achieved by wavelet was 90.07%. Curvelet transform achieved 94.28% for abnormal class as compared to wavelet functions that achieved 83%. Eltoukhy et al. [10] Used curvelet transform in feature extraction. The coefficients of the object are used as feature vector. The Euclidian distance is used in classification technique. Classification accuracy rate gave a 98.59%.

Gardezi et al. [11] Used local discrete cosine transforms (LDCT) and curvelet transform via wrapping technique. The statistical features includes energy, contrast, entropy, correlation and inverse difference moment are extracted. Those features extracted by LDCT in the mammogram images. The detection rate is reached to 77.3%. Eltoukhy et al. [12] [13] proposed a method for extracting the most

*Corresponding author e-mail: Mohamed_eltokhy@ci.suez.edu.eg

significant features from mammogram breast cancer images. The proposed method used multiresolution representation. Coefficients are obtained to construct the matrix $K \times N$, where K is the number of images and N is the number of coefficient for each images. The features are ranked in descending order and features capabilities are calculated. Support vector machine used in classification step. The classification accuracy rate achieved 95.84% using wavelet coefficients for normal versus abnormal using 1238 coefficients. The classification accuracy rate is 96.56% to determine if a tumor is benign and malignant using only 150 features.

Shareef [14] used the algorithm of marker controlled watershed transformation. It started by filter the image and compute watershed area on gradient magnitude. Then compute foreground markers. Then compute background markers. Then compute watershed area. Finally mark the object on the original image and color assign to each object based on the number of objects in the label matrix and range of colors in the color map and the obtained result was 84.848%.

From the literature, we can notice the importance of segmentation and feature extraction steps to develop an efficient computer aided diagnosis system. There are some limitations of the developed system such as over segmentation in finding the region of interest, identifying the best set of feature that able to discriminate between different classes normal or abnormal and benign or malignant. In this paper, the main objective is to develop a computer aided system for breast cancer detection and diagnosis in ultrasound images. The system starts with segmentation step. This step is used to separate the object from its background. It is the primary step to the feature extraction and classification steps. The aim of segmentation method is to solve the problem of over segmentation. It is then followed by feature extraction step, the aim is to find the most significant features that able to discriminate normal and focal lesions and then classify focal lesions as benign or malignant tumors. Finally proposed method compare between three types of classifiers SVM, KNN and CART classifiers.

2 Methodology

In this section, an overview about the collected ultrasound data is presented. Then, it is followed by description of the proposed system steps.

2.1 Image dataset

In this study the dataset is consist of 91 breast ultrasound images. It was collected from Damietta ecology institute. This dataset has been taken by ultrasound device called logic p5 high resolution superficialx probe 7.5 MHZ. it consists of 36 malignant images, 44 benign images and 11 normal images.

2.2 Image segmentation

Image segmentation is an important step to detect the region of interest (ROI) [14]. It can subtract the object (abnormal region) from the background. Marker controlled watershed transformation used for segmentation step.

Watershed transformation: One can imagine a landscape or topographic relief being emerged in a lake. With holes pierced in local minima, basins which will be filled up with water starting at these local minima. At points where water coming from different basins would meet, dams are built when the water level has reached the highest peak in the landscapes. As a result, the landscape is divided into regions or basins which are separated by dams that are called watershed lines or simply watershed. Watershed transformation has a big problem namely over segmentation [14]. It can produce more than one region as the output. This problem comes mostly from the noise. Automatic markers in the desired areas to be segmented are put. This method controls the flooding only to the catchment basins which are associated with each marker. This method called marker-controlled watershed transformation.

Proposed method:

Modified marker controlled watershed transformation are presented as follows:

1) Compute segmentation function.
2) Compute the pixels that belong to the object.
3) Compute watershed area.
4) Draw a contour plot to mark the object in the images.

The proposed method starts firstly reading images and converts them into grayscale.

1) It computes segmentation function by computing gradient magnitude and segments the image using the watershed transform directly on the gradient magnitude.
2) The proposed method computes the foreground marker that marks the abnormal regions. It marks pixels within the objects using morphological structuring element that creates a structuring element of the type specified by shape. The proposed method uses the ellipse shape. It is the most appropriate and suitable for the earliest form of tumor. Then uses morphological technique called opening-by-construction and closing-by-construction. Finally computes regional minima of image. We noted that over segmentation problem has been solved after this step as shown in fig. 1.

Fig. 1: (a) Image with over segmentation, (b) image after solving over segmentation.

3) Computes watershed area after solving the problem of over segmentation.
4) The fourth step marks the object i.e. mark the abnormal region. The proposed method draw a contour plot from the object that obtained in step 2. The coordinates are obtained to mark the abnormal region in the original images. Fig.2 illustrates the proposed method steps.

2.3 Feature extraction

The proposed method compute features from the segmented regions. Feature extraction is an important step in the diagnosis process [15]. Features are extracted to classify the breast ultrasound images as normal or focal lesion and then to classify focal lesions into benign or malignant. The proposed method used texture features including homogeneity, contrast, energy, correlation and entropy [16] [17] [18]. Statistical features include mean, standard deviation, variance and median. Those features are calculated onto four wavelet decomposition matrices [9] [12]. Four different decomposition levels based on Daubechies-4 (db4) wavelet function are used [9]. 144 coefficients are obtained as a feature vector, i.e. each ultrasound image is represented by 144 coefficients. These coefficients are presented to the classifier in the classification step.

2.4 Classification

KNN, SVM and CART are used in the proposed method. Those classification techniques are used to differentiate normal or focal lesion and classify focal lesion as benign or malignant. KNN is the most common classification techniques which play an important role due to its simplicity [19]. Training data set is used which contains sample and target variables. The sample variables are compared with the distance of the unknown to KNN to determine its class by averaging the class numbers of K nearest points or by a majority vote that obtained for them. SVM is a successful method in the classification step [20] [21].

The main idea in SVM is to make a hyper plane in an infinite and a high dimensional space [22]. CART [23] is used to predict the response to data. Classification trees are used to identify the class to which the data belongs. Regression trees are used to predict the value of the target variable.

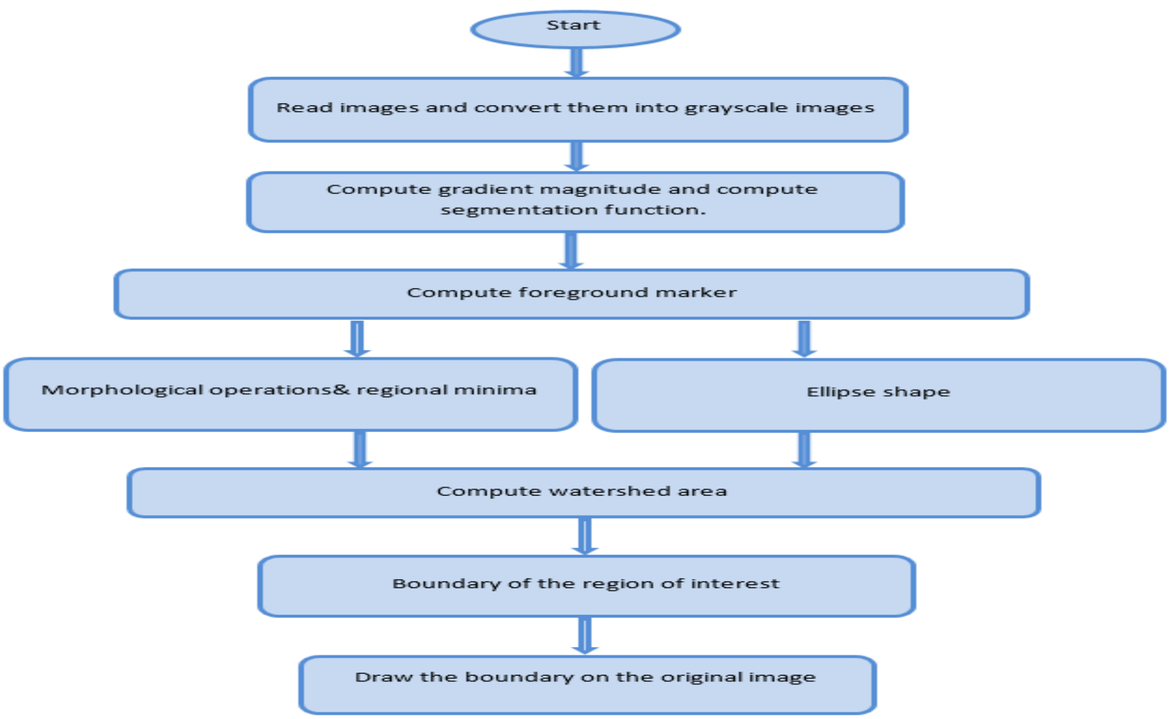

Fig. 2: Flow chart of proposed method.

3 Results and discussion

3.1 Result of segmentation technique

The work in this paper is divided into two phases. In phase one, the system starts to find the region of interest, we call it segmentation step. This step is used to detect the suspicious region. In the second phase, the aim is to determine whether the detected region is belongs to the normal or abnormal regions then the abnormal region classified as benign or malignant. The results of the first phase illustrated as below.

Table 1: The results of proposed segmentation method.

Item	Proposed method
Sensitivity	91.25%
Specificity	100%
Accuracy	92.31%
PPV	100%
NPV	61.11%

The performance evaluation consists of the relevance between results of segmentation and the images from the expert. The proposed method results are matched with the marked region from the expert i.e. the abnormal region that the specialist determined in images matches the abnormal region that determined using the proposed system. The reference for the final diagnosis depends on US and mammographic diagnosis as well as biopsy. The proposed system is applied over the dataset. The dataset containing 91 breast ultrasound images, 44 images are benign, 36 images are malignant and 11 images are normal. The performance evaluation metrics sensitivity, specificity, accuracy, negative predictive value (NPV) and positive predictive value (PPV) are calculated using the following formulas. Table 1 illustrates the segmentation accuracy rates. Fig 3 shows the results of the segmentation technique step by step.

The comparison between the obtained results and the results in Ref [14] shows the improvement of accuracy to become 92.31% and reduced the time to become 4 second instead of 84.848% with 7 second for each one image.

$$\text{Sensitivity} = \frac{TP}{TP+FN} x100$$

$$\text{Specificity} = \frac{TN}{TN+FP} x100$$

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} x100$$

$$\text{NPV} = \frac{TN}{TN+FN} x100$$

$$\text{PPV} = \frac{TP}{TP+FP} x100$$

Where, TP: is True Positive (sick people diagnosed correctly as sick people). TN: is True Negative (health people diagnosed correctly as health people). FP: is False Positive (health people diagnosed incorrectly as sick people). FN is False Negative (sick people diagnosed incorrectly as health people).

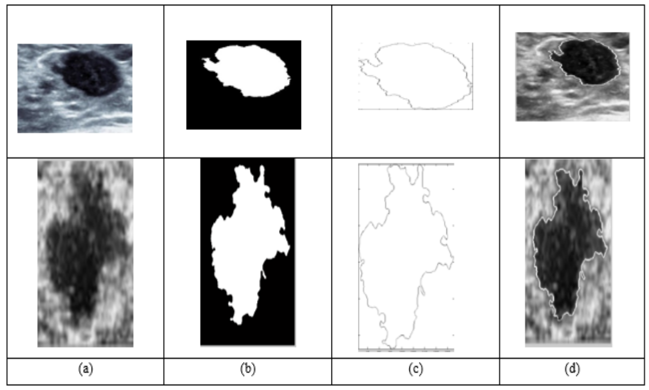

Fig. 3: Illustrate the results of the segmentation technique, Images in column (a) The original US images. (b) The results of step 2 after modification of the over segmentation. (c) Contour coordinates which obtain from images in step2. (d) The results of the marker function.

3.2 Result of classification step

Classification is an important step in detecting breast cancer. Features are extracted to classify different classes. In the case of classification between normal and abnormal classes, Table2 Shows the result of accuracy, sensitivity, specificity, PPV and NPV using KNN, SVM and CART classifiers. SVM and CART classifiers gave the highest accuracy using texture features to become 100%, while in case of KNN classifier, the classification rate using texture features decreased to become 87.91 %. *i.e.* the texture feature is the most significant features with the best accuracy to differentiate between normal and abnormal classes. SVM and CART classifiers gave the same accuracy using statistical and texture features to become 98.90%, while KNN classifier decreased to become 87.91%. KNN achieved 98.90% classification rate using statistical features, while SVM and CART classifiers achieved 86.81% and 97.80%, respectively.

Fig. 4, illustrates the results of classification rate with SVM, CART and KNN classifiers using texture feature when differentiate between normal and abnormal classes. CART and SVM classifiers achieve the best results.

In the case of classifying between benign and malignant classes, Table 3 shows the results of accuracy, sensitivity, specificity, PPV and NPV. CART classifier gave the best classification rate using statistical and texture features to become 83.75%, while it decreased to 81.25% and 73.75% using SVM and KNN, respectively. KNN classifier achieved 81.25% classification rate using statistical feature, while it decreased to 65% and 72.5% using SVM and CART classifiers, respectively. All classifiers achieved the same classification rate using texture feature.

Fig. 5, illustrates the results of classification rate with SVM, CART and KNN classifiers using texture and statistical features when differentiate between benign and malignant tumor. CART classifier achieves the best result.

Table 2: The accuracy, sensitivity, specificity, PPV and NPV to differentiate between normal and abnormal.

Features	Performance	SVM	KNN	CART
SF&TF	Accuracy	98.90%	87.91 %	98.90 %
	Sensitivity	98.70%	100%	98.75%
	Specificity	100%	100%	100%
	NPV	91.67%	100%	91.67%
	PPV	100%	100%	100%
SF	Accuracy	86.81%	98.90 %	97.80 %
	Sensitivity	98.75%	98.7%	97.50%
	Specificity	0%	100%	100%
	NPV	0%	91.67%	84.20%
	PPV	87.78%	100%	100%
TF	Accuracy	100 %	87.91 %	100%
	sensitivity	100%	100%	100%
	Specificity	100%	100%	100%
	NPV	100%	100%	100%
	PPV	100%	100%	100%

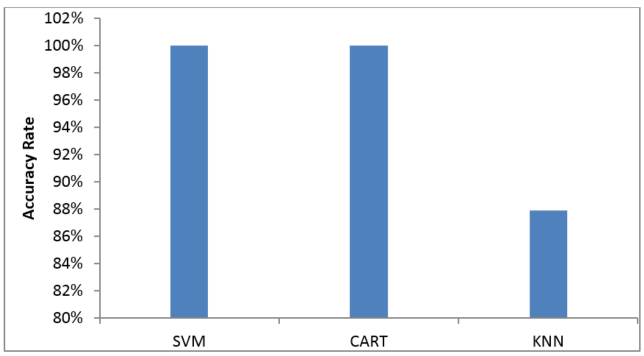

Fig. 4: The classification accuracy of SVM, CART and KNN using texture feature.

Table 3: Show the accuracy, sensitivity, specificity, PPV and NPV to classify between benign and malignant.

Features	Performance	SVM	KNN	CART
SF&TF	accuracy	81.25%	73.75%	83.75 %
	sensitivity	80.56%	63.89%	80.56%
	specificity	81.81%	81.81%	86.36%
	PPV	78.38%	74.19%	82.86%
	NPV	83.72%	73.45%	84.44%
SF	accuracy	65.00%	81.25%	72.50%
	sensitivity	41.67%	80.56%	66.67%
	specificity	84.09%	81.81%	77.27%
	PPV	68.18%	78.38%	70.59%
	NPV	63.795	83.72%	73.91%
TF	accuracy	78.75%	78.75%	78.75%
	Sensitivity	75.00%	72.22%	69.44%
	specificity	81.81%	84.09%	86.36%
	PPV	77.14%	78.79%	80.65%
	NPV	80.00%	78.72%	77.55%

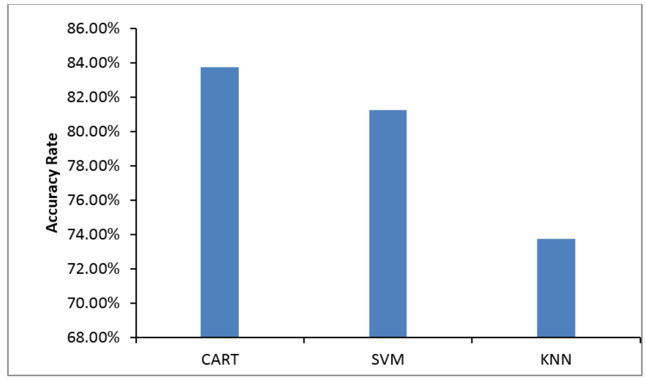

Fig. 5: The performance of SVM, CART and KNN using both texture and statistical features.

4 Conclusion

The proposed method applied a system for detection and diagnosis of breast cancer in ultrasound images. Marker controlled watershed transformation technique is applied in the segmentation step to detect the region of interest (cancer region). The problem of over segmentation has been solved by using the ellipse shape which is the most appropriate and suitable for the earliest form of tumor. The Proposed method extracts the features from the segmented region. The used texture and statistical features are based on wavelet decomposition. KNN, SVM and CART classifiers are used to classify region of interest as normal or focal lesion and then classify focal lesion as benign or malignant.

The proposed system has been evaluated by calculating the performance evaluation metrics (sensitivity, specificity, PPV, NPV and accuracy). The obtained results demonstrate that the proposed method could contribute to the successful detection of breast cancer in ultrasound images. As well as, SVM and CART classifiers gave the best results in case of differentiating between normal and focal lesion using texture feature. It also shows that, the CART classifier gave the best result in case of differentiating between benign and malignant using both texture and statistical features.

References

[1] R. Siegel, D. Naishadham, A. Jemal, *Cancer statistics 2013, CA: A cancer journal for clinicians*, vol. **63**, no. 1, pp. 11-30, (2013).

[2] W. Gomez, L. Leija, W. C. A. Pereira, A. Alvarenga, A. F. C. Infantosi, Computerized lesion segmentation of breast ultrasound based on marker−controlled watershed transformation, *Medical Physics*, vol. **37**, no. 1, pp. 82-95, (2010).

[3] P. Telagarapu and S. Poonguzhali, Analysis of contourlet texture feature extraction to classify the benign and malignant tumors from breast ultrasound images, *International Journal Of Engineering And Technology*, vol. **6**, no. 1, pp. 293-305, (2014).

[4] H. D. Cheng, S. Juan, J. Wen, G. Yanhui, and Z. Ling, Automated breast cancer detection and classification using ultrasound images: A survey. *Pattern Recognition*, vol. **43**, issue 1, pp. 299-317, (2010).

[5] P. Crystal, S.D. Strano, S. Shcharynski, and M.J. Koretz, Using sonography to screen women with mammographically dense breasts, AJR. *Am. J. Roentgenol*, vol. **181**, no. 1, pp. 177-182, (2003).

[6] P. Skaane, Ultrasonography as adjunct to mammography in the evaluation of breast tumors, *Acta Radiol Suppl*, vol. **420**, pp. 1-47, (1999).

[7] A.T. Stavros, D. Thickman, C.L. Rapp, M.A. Dennis, S.H. Parker, and G.A.Sisney, Solid breast nodules use of sonography to distinguish between benign and malignant lesions, *Radiology*, vol. **196**, no. 1, pp. 123-134, (1995).

[8] H.M. Zonderland, E.G. Coerkamp, M.J. devijver, and A.E. voorthuisen, Diagnosis of breast cancer: contribution of US as an adjunct to mammography, *Radiology*, Vol. **213**, No. 2, pp. 413-422, (1999).

[9] M.M. Eltoukhy, I. Faye, B.S. Belhaouari, A comparison of wavelet and curvelet for breast cancer diagnosis in digital mammogram, *Computers In Biology And Medicine*, vol. **40**, issue 4, pp. 384-391, (2010).

[10] M.M. Eltoukhy, I. Faye, B.S. Belhaouari, Breast cancer diagnosis in digital mammogram using multiscale curvelet transform, *Computerized Medical Imaging And Graphics*, vol. **34**, issue 4, PP. 269-276, (2010).

[11] S. J. safdar Gardezi, I. Faye, N. Kamel, M. M. Eltoukhy and M. Hussain, Performance analysis of morphological component analysis (MCA) method for mammograms using some statistical features, *AIP Conference Proceedings*, Vol. **1621**, Issue 1, PP. 416-421, (2014).

[12] M. M. Eltoukhy, I. Faye, S. B. Brahim, A statistical based feature extraction method for breast cancer diagnosis in digital mammogram using multiresolution representation, *Computers In Biology And Medicine*, Vol **42**, No. 1, pp. 123-128, (2012).

[13] M. M. Eltoukhy, I. Faye, An optimized feature selection method for breast cancer diagnosis in digital mammogram using multiresolution representation, *Appl. Math. Inf. Sci.* Vol. **8**, no. 6, PP. 2921-2928, (2014).

[14] S. R. Shareef, Breast cancer detection based on watershed transformation, *International Journal Of Computer Science*, Vol. **11**, issue 1, no 1, pp. 237-245, (2014).

[15] Y. Y. Tang, Y. Tao and E. C. M. Lam, New Method for Feature Extraction based on Fractal Behavior, *Pattern Recognition*, Vol. **35**, No. 5,PP. 1071-1081, (2002).

[16] A. K, Dr. K. Sankaranarayanan, statistical feature extraction to classify oral cancers, *Journal of Global Research in Computer Science*, Volume **4**, No. 2, (2013).

[17] R.nithya, B.santhi, comparative study on feature extraction method for breast cancer classification, *Journal of Theoretical and Applied Information Technology*, Vol. **33**, No.2, (2011).

[18] M. Radhakrishnan and T. Kuttiannan, comparative analysis of feature extraction methods for the classification of prostate cancer from TRUS medical images, *IJCSI International Journal of Computer Science Issues*, Vol. **9**, Issue. 1, No. 2, (2012).

[19] R. J. Ramteke, K. Monali, Automatic medical image classification and abnormality detection using K- nearest neighbour, *International Journal of Advanced Computer Research*, Vol. **2**, No. 4, Issue. 6, (2012).

[20] Y. L. Huang, K. L. Wang and D.R. Chen, Diagnosis of breast tumors with ultrasonic texture analysis using support vector machines, *Neural Computing and Applications*, Vol. **15**, issue. 2, PP. 164–169, (2006).

[21] K. Kwang, J. Keechul, P. H. Se and K. J. Hang, Support vector machines for texture classification, *IEEE Transactions Pattern Analysis And Machine Intelligence*, Vol. **24**. No. 11, pp. 1542–1550, (2002).

[22] J. S. Taylor and N.Cristianini, An introduction to support vector machines and other kernel-based learning methods, Cambridge University Press New York, NY, USA, (2000).

[23] B. Leo, F. Jerome, O. Richard and S. Charles, classification and regression trees, CRC press, (1984).