

## A Novel Pectoral Removal Approach in Digital Mammogram Segmentation and USRR Feature Selection

R.Poongothai\*, I.Laurence Aroquiaraj\*\*, D.Kalaivani\*\*\*

\* (Research Scholar, Periyar University, Salem)

\*\* (Assistant Professor, Periyar University, Salem)

\*\*\* (Research Scholar, Periyar University, Salem)

### Abstract

Detection of micro calcification based on textural image segmentation and classification is the most effective early-diagnosis of breast cancer. The aim of segmentation is to extract a breast region by estimation of a breast skin-line and a pectoral muscle as well as removing radiographic artifacts and the background of the mammogram. The intensity value is taken using histogram function. After detecting the region intensity value from mammogram, the edge between pectoral region and breast region will be detected using Fuzzy Connected Component Labeling. Our proposed method worked good for separating pectoral region by eliminating cancer area. The raster scan method is used for fixing the pectoral removal area in the original image. After removal pectoral muscle from the mammogram, so that further processing is confined to the breast region alone. Two error measures were used to compare these three methods performance. One of the measures is Mean absolute error (MAE) and another one is hausdroff distance measure to find the distance between the binary pectoral region and binary breast region alone. We applied unsupervised feature selection of rough set based Relative Reduct algorithms and it demonstrates effectively remove the redundant features. The quality of the reduced data is measured by the classification performance and it is evaluated using WEKA classifier tool.

**Keywords:** Digital mammogram, morphological reconstruction, Fuzzy connected component labeling, segmentation, feature selection.

### I. Introduction

Breast cancer is the second most major health problem in developed countries. According to the latest study from NCI (U.S. National Cancer Institute) more than 207,090 new cases were reported and about 39,840 were the death statistics in 2010. Nearly 1.4 million US women have a family history of breast cancer. As there is no effective method for its prevention, the diagnosis of breast cancer in its early stage of development has become Very crucial for the prevention of cancer. Computer-aided diagnosis (CAD) systems play an important role in earlier diagnosis of breast cancer [1].

Another method such as a biopsy is used for detection of breast cancer, where the patient undergoes a surgical procedure. Many CAD based systems were used by the radiologist for the diagnosis of breast masses. The limitation of these systems is mainly at breast mass detection. Therefore, accurate segmentation of a breast mass is an important step for the diagnosis of breast cancer in mammography. Breast mass is a localized swelling, which is described by its characteristics. Usually these boundaries are of varying size and shape. Because of this, breast mass segmentation is a challenging task.

The rest of the paper is organized as follows. Section 2 gives an introduction to the previous work. Section 3 present preprocessing using Median Filter. Section 4 presents pectoral muscle removal using Fuzzy with CCL. Section 5 presents Segmentation using Watershed Transformation. Section 6 describes the GLCM for Feature Extraction. Section 7 presents the proposed unsupervised relative reduct (USRR) algorithm for feature selection. Section 8 describes the classifier tool. The experimental results are discussed in Section 9 and conclusion is presented in Section 10.

### II. Previous Work

In the literature, several methods have been proposed to identify the pectoral muscle in mammograms. T.S.Subashini et al. proposed a technique for pectoral muscle removal based on connected component labeling. To extract the pectoral muscles from this image, binary image should be obtained [4]. Sara Dehghani et al employed a technique for removal of dark background parts which are not important in processing of mammography images [3]. Roshan Dharshana Yapa proposed a technique to compare performance of connected component labeling algorithms on grayscale digital mammograms [8]. In Ali Cherif Chaabani et al. Proposed to extract the breast region, we used a method based on automatic thresholding (Otsu's) and connected component labeling algorithm. Identifying the pectoral muscle has been done using Hough transform and active contour [7]. Masek et al [9] employed both a threshold-based algorithm and a straight line fitting

technique to represent the pectoral muscle. The straight line representation by this method is not an efficient way to identify the pectoral muscle due to the existence of curved pectoral muscles.

### III. Preprocessing

Mammograms are difficult images to interpret, and a preprocessing phase is necessary to improve the quality of the images and make the feature extraction phase more reliable. In order to limit the search for abnormalities by computer aided diagnosis systems to the region of the breast without undue influence from the background of the mammogram, removal of artifacts and removal of pectoral muscle is necessary. Preprocessing stage consists of two parts. The first part involves the removal of unwanted parts from the image and the second part deals with reducing the high frequency components present in the image. Artifacts are removed by morphological open operation followed by reconstruction operation [2]. Median filtering is similar to using an averaging filter, in that each pixel is set to an average of the pixel values in the neighborhood of the corresponding input pixel median filtering is better able to remove these outliers without reducing the sharpness of the image.

### IV. Pectoral Muscle Detection and Removal

Pectoral muscles are the regions in mammograms that contain brightest pixels. These regions must be removed before detecting the tumor cells so that mass detection can be done efficiently. Pectoral muscles lie on the left or right top corner depending on the view of the image. We must detect the position of the pectoral muscles (left top corner or right top corner) as is shown in the figure 1(a) and (b), before removing it. For this searching for nonzero pixels are simultaneously done from the left and right top corner. Width of the image in which the non zero pixel detected from both the corner were counted and compared. If the left width is smaller than the right width then it is assumed that pectoral is on the left side of the image else it is on the right side.

From the detected corner pixel the intensity discontinuity is detected on each and every column of the same row. Coordinates of the pixel in which the intensity change is encountered is considered as width of the pectoral region. All the pixels, which lie inside pectoral width and half of the height of the whole image is segmented from the original image. This rectangle shaped image contains the entire pectoral muscles.

To extract the pectoral muscles [18] from this image binary image should be obtained by simple thresholding. This binary image contains pectoral muscles and other tissues. To segment the pectoral muscles alone from the binary image raster

scanning is done from the right or left side of the image to detect the intensity discontinuities. The resulting image contains pectoral muscles alone and this region is completely removed from the original image.

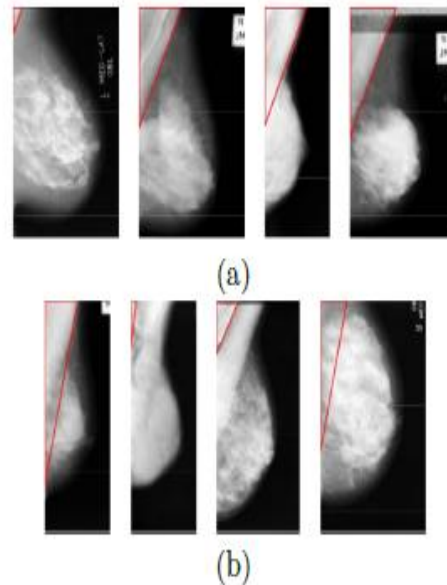


Figure 1: (a) and (b): Different sizes and intensities of the pectoral muscle in breast mammograms.

### Procedure: merging Fuzzy with CCL algorithm

INPUT: over segmented image

OUTPUT: Pectoral muscle removal from the Original image

1. Iterate through each element of the data by column, then by row (Raster scanning).
2. If the element is not the background. Get the neighboring elements of the current element. If there is no neighbors, uniquely label the current element and continue.
3. Otherwise, find the neighbors with the smallest label and assign it to the current element. Store the equivalence between neighboring labels.
4. Iterate through each element of the data by column, then by row.
5. If the element is not the background. Re-label the element with the lowest equivalent label.
6. To fuzzy construction, the input grays is ranged from 0-255 gray intensity, and according to the desired rules the gray level is converted to the values of the membership functions.
7. Compare the cropped image and original image. Then retrieving pixels from the original image.

### V. Segmentation

Segmentation [19] refers to the process of partitioning a digital image into multiple segments



formed by the union of connected pixels. The goal of segmentation is to simplify the representation of an image into something that is more meaningful and easier to analyze, regions. Image segmentation is typically used to locate objects and boundaries. More precisely, image segmentation [20] is the process of assigning a label to every pixel with the same visual characteristics such as color, intensity, or texture. The result of image segmentation is a set of regions that collectively cover the entire image, or a set of contours extracted from the image. After median filter, Watershed Transformation Segmentation (WTS) is used to identify the suspicious region or micro-calcification. An example of image segmentation for the approach is given in Figure 2.

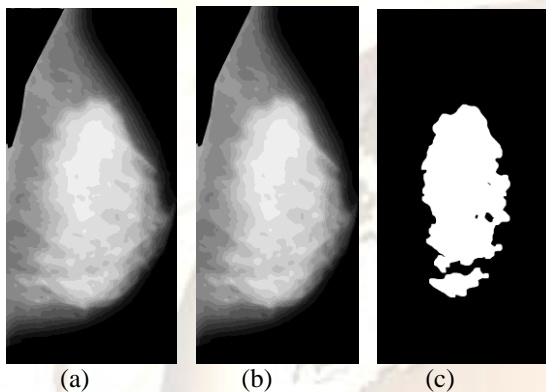


Figure 2. Example for Watershed Transformation Segmentation in the image mdb003.pgm. (a) Original image, (b) Preprocessing image, (c) Process of Watershed segmentation method.

## VI. Feature Extraction

The texture coarseness or fineness of an image can be interpreted as the distribution of the elements in the description matrix. The textural description matrices GLCM are created for each segmented mammogram image for defined distance and direction. The Haralick features are extracted from the textural description matrices. A feature value [17] is a real number which encodes some discriminatory information about a property of an object. It may not always be obvious what type of information or feature is useful for a particular detection task. The texture analysis matrix itself does not directly provide a single feature that may be used for texture discrimination. Instead, the matrix can be used as a representation scheme for the texture image and the features are computed. The features based on the distribution matrices should therefore capture some characteristics of textures such as homogeneity, coarseness, periodicity and others.

### A. Gray-Level Co-occurrence Matrix Features

It is a statistical method that considers the spatial relationship of pixels is the gray-level co-occurrence matrix (GLCM), also known as the gray-

level spatial dependence matrix [13]. By default, the spatial relationship is defined as the pixel of interest and the pixel to its immediate right (horizontally adjacent), but you can specify other spatial relationships between the two pixels. Each element  $(I, J)$  in the resultant GLCM is simply the sum of the number of times that the pixel with value  $I$  occurred in the specified spatial relationship to a pixel with value  $J$  in the input image

The Following GLCM features were extracted in our research work [10]: Angular Second Moment (F1), Contrast (F2), Correlation (F3), Sum Of Squares: Variance (F4), Inverse Difference Moment (F5), Sum Average(F6), Sum Variance(F7), Sum Entropy(F8), Entropy(F9), Difference Variance(F10), Difference Entropy(F11), Information Measures Of Correlation-I(F12), Information Measures Of Correlation-II(F13), The Maximal Correlation Coefficient(F14).

## VII. Feature Selection using Unsupervised Relative Reduct (USRR)

In [11] [14], a FS method based on a relative dependency measure was presented. The technique was originally proposed to avoid the calculation of discernability functions or positive regions, which can be computationally expensive without optimizations. The authors replaced the traditional rough set degree of dependency with an alternative measure, the relative dependency. The USRR algorithm 1 is given below.

Algorithm 1. Shows an Unsupervised Relative Reduct

## VIII. Classifications

The classifier tool WEKA [12] is open source java based machine-learning workbench that can be run on any computer in which a java run time

Relative reduct  $(C, D)$   
 $C$ , the set of all conditional features;  
 $D$ , the set of decision features;  
 (1)  $R \leftarrow C$   
 (2)  $\forall a \in C$   
 (3) if  $(\kappa_{R - \{a\}}(D) = 1)$   
 (4)  $R \leftarrow R - \{a\}$   
 (5) return  $R$

environment is installed. It brings together many machine learning algorithm and tools under a common frame work. The WEKA is a well known package of data mining tools which provides a variety of known, well maintained classifying algorithms. This allows us to do experiments with several kinds of classifiers quickly and easily. The tool is used to perform benchmark experiment.

**IX.Results and Discussions**

The proposed algorithm is simulated in MATLAB and applied on several test images. The results of Mean Absolute Error (MAE) and Hausdroff Distance Measures for all the proposed two methods were compared with each other. The pectoral muscle region removal from the mammogram is had by our three proposed method. Two methods were CCL(connected component labeling)+Fuzzy method, Straightline with Fuzzy. These two methods were compared with each other by error measure and distance measure. The Mean Absolute Error is calculated for some 14 mammogram images and it is shown in given figure (3). The Hausdroff Distance is calculated for same 14 Pectoral removal mammogram images and it is shown in figure(4). For the total number of MIAS images, pectoral removed images were detected with each method and the average value is calculated for MAE and hausdroff distance which in the Table(1). Then segmentation can be performed using watershed segmentation transformation and 14 Haralick features are extracted using gray-level co-occurrence matrix (GLCM). The best features can be selected based on relative dependency measure that is Unsupervised Relative Reduct (RR) method. The selected features are classified by using WEKA classification. It is found that the classification accuracy increased and decreased mean absolute error, which is shown in figure (5) and (6). The classification results, clearly indicates that the method selects useful features which are of comparable quality.

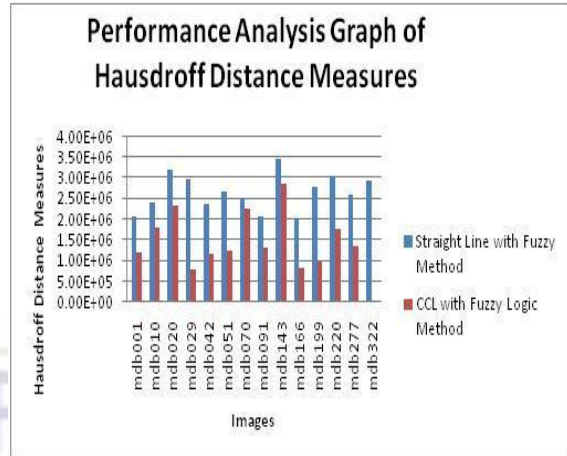


Figure (4) shows Performance Analysis of Hausdroff Distance Measures

Total number of Mammogram images taken (MIAS Database)	Average for Straight line with Fuzzy		Average for CCL with Fuzzy	
	MAE	H.D	MAE	H.D
322	0.7198	2652.30	0.5164	1670.528

Table (1) shows Performance Analysis for the average of MAE and Hausdroff Distance Measures

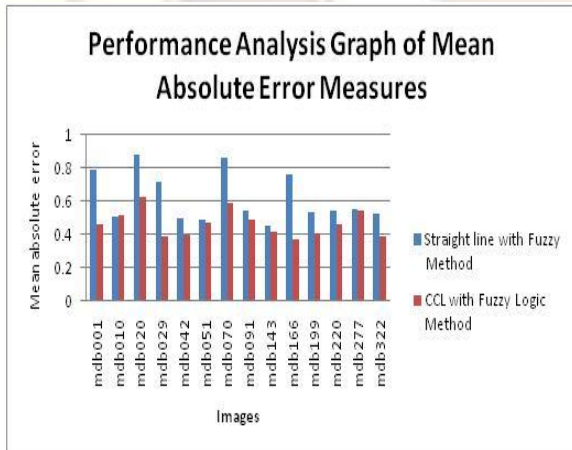


Figure (3) shows Performance Analysis of Mean Absolute Error Measures

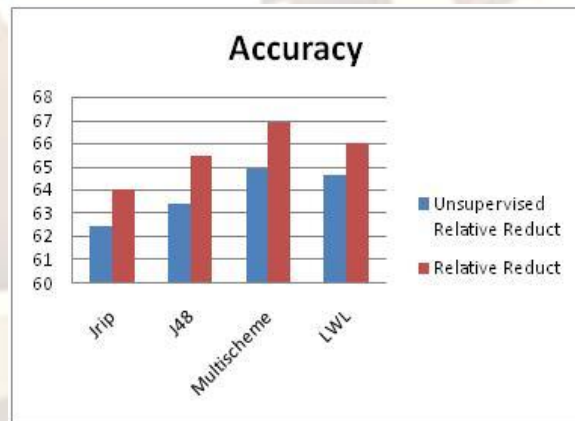


Figure (5) shows Accuracy of Classification.

**X.Conclusion**

This paper presents a method to detect pectoral muscle and segment the masses from the mammography. The proposed work was done using Mini-Mias mammogram database. Here we have presented several aspects of image processing techniques that can be applied for detection of pectoral muscle in digital mammography. To removing background noise from the original image



we used median filter, then pectoral muscle removal using Fuzzy with connected component labeling.

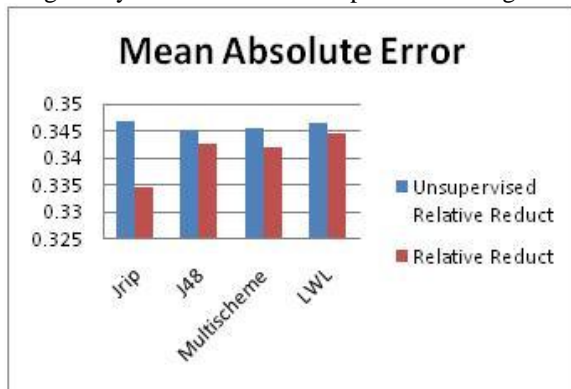


Figure (6) shows Mean Absolute Error Value.

In this paper, we have considered the problem of detecting cancer masses by the application of simple thresholding followed by connected component labeling with fuzzy and an algorithm to remove artifacts in digital mammograms using morphological open operation followed by reconstruction. Further the pectoral muscle was removed successfully using simple thresholding and raster scan methods. Then features selected based on relative reduct method and classification using WEKA classifier. Finally, as a future work we will consider more samples and mass shape spectral identifications in mammograms using near sets.

## References

- [1] S. Don, Eumin Choi, and Dugki Min, Breast Mass “Segmentation in Digital Mammography Using Graph Cuts” *ICHIT 2011*, CCIS 206, pp. 88–96, Springer-Verlag Berlin Heidelberg 2011.
- [2] Aswini Kumar Mohanty, Mrs. Swasati Sahoo, Mrs. Arati Pradhan, Saroj Kumar Lenka, “Detection of Masses from Mammograms Using Mass shape Pattern”, Vol 2 (4), pp.1131-1139, ISSN:2229-6093, *IJCTA* July-August 2011.
- [3] Sara Dehghani and Mashallah Abbasi Dezfooli, “A Method for Improve Preprocessing Images Mammography” *International Journal of Information and Education technology*, Vol.1, No.1, April 2011 ISSN: 2010-3689.
- [4] T.S.Subashini, V.Ramalingam, S.Palanivel, “Pectoral muscle removal and Detection of masses in Digital Mammogram using CCL”, *International Journal of computer Applications* (0975-8887) Volume 1 – No.6,2010.
- [5] Smith, R.A., Cokkinides, V., Brawley, O.W.: Cancer Screening in United State, 2010: “A review of current American Cancer Society”, Guidelines and Issues in Cancer Screening. CA Cnacer J. Clin. 60, 99–119, 2010.
- [6] Lee, C.-S., Kanmaz, T.J.: “Recent Controversies in Cancer Screening Recommendations” *US Pharm.*, vol. 35(11), pp. 3–8 (2010) (oncology Supply).
- [7] In 2010, Ali Cherif Chaabani, Atef Boujelben, Adel Mahfoudhi, Mohamed Abid, *International Journal of Digital Content Technology and its Applications*, Volume 4, Number 3, June 2010.
- [8] Roshan Dharshana Yapa and Koichi Harada “Connected Component Labeling Algorithms for Gray-Scale Images and Evaluation of Performance using Digital Mammograms”, Department of Information Engineering, Hiroshima University, Japan. *International Journal of Computer Science and Network Security*, Vol.8 No.6, June 2008.
- [9] Suckling J, Parker J, Dance D, Astley S, Hutt I, Boggis C, et al. “The mammographic images analysis society digital mammogram database”, 2<sup>nd</sup> *International Workshop on Digital Mammography*, York, England, *Experpta Med* 1994;1069:pp375-378.
- [10] R. M. Haralick, K. Shanmugan, I. Dinstein, “Textural features for image classification”, *IEEE Trans. Syst., Man, Cybern.*, Vol. 3(6), 1973, pp. 610–621.
- [11] J. Han, X. Hu, and T.-Y. Lin, “Feature subset selection based on relative dependency between attributes,” in Proc. of the 4th *International Conf. on Rough Sets and Current Trends in Computing*, Uppsala, pp. 176–185, 2004.
- [12] Holmes, G., Donkin, A., Witten, I.H.: “WEKA: a machine learning workbench.” In: Proceedings Second Australia and New Zealand Conference on *Intelligent Information Systems*, Brisbane, Australia, pp. 357-361 (1994).
- [13] M. Vasantha “Medical Image Feature, Extraction, Selection And Classification” *European Journal of Scientific Research* Vol. 2(6), 2071-2076, ISSN 1450-216X, Vol.50 No.4 (2011), pp.543-553,2010.
- [14] R. Roselin, K. Thangavel, C. Velayutham, “Fuzzy Rough Feature Selection for Mammogram Classification”, *International Journal of Electronic Science and Technology (JEST)*, Vol. 9, No. 2, pp 124-132, June 2011.
- [15] K.Thangavel, M.Karnan, R.Sivakumar, A. KajaMohideen “Ant Colony System for Segmentation and Classification of Microcalcification in Mammograms”, *International Journal on Graphics, Vision and Image Processing*, GVIP, P1150619001, 07, Special Issues on Mammograms, ISSN :

- 1687-398X-1687-3998, pp. 1-12, 2005.
- [16]. M. A. Duarte, A. V. Alvarenga, C. M. Azevedo, A. F. C. Infantosi, W. C. A. Pereira "Automatic Microcalcifications Segmentation Procedure Based on Otsu's Method and Morphological Filters", RIO DE JANEIRO, BRAZIL IEEE 102 ISBN: 978-1-61284-918-8 Vol. 11, pp. 103-108, IEEE Catalog Number: CFP1118G-ART 2011.
- [17]. Shu-Ting Luo & Bor-Wen Cheng "Diagnosing Breast Masses in Digital Mammography Using Feature Selection and Ensemble Methods" J Med Syst DOI 10.1007/s10916-010-9518-8 © Springer Science & Business Media, LLC, Vol. 34, 2010.
- [18]. Dharendra Pal Singh and Ashish Khare, "Automatic Image Segmentation using Threshold Based Parameter and Genetic Algorithm", *International Journal of Research and Reviews in Computer Science (IJRRCS)*, Vol. 3, No. 3, pp. 1672-1677, ISSN: 2079-2557 © Science Academy Publisher, United Kingdom, June 2012.
- [19]. Guillaume Kom, Alain Tiedeu, Martin Kom, John Ngundam, "a comparative evaluation of two algorithms of detection of masses on mammograms", *Signal & Image Processing: An International Journal (SIPIJ)*, Vol.3, No.1, February 2012.
- [20]. C. Karthikeyan, Dr. B. Ramadoss, "Image Registration and Segmentation for the Quantitative Evaluation of Complex Medical Imaging and Diagnosis over Head – A Systematic Review" *International Journal Comp. Tech. Appl.*, Vol 2 (4), 1058-1061 IJCTA, pp.1 1058-1061, ISSN:2229-6093, JULY-AUGUST 2011.
- [21]. Jelena Bozek, Mario Mustra, Kresimir Delac, and Mislav Grgic "A Survey of Image Processing Algorithms in Digital Mammography" M. Grgic et al. (Eds.): Rec. Advan. in Mult. Sig. Process. and Commun., SCI 231, pp. 631–657, © Springer-Verlag Berlin Heidelberg 2009.
- [22]. Vibha S. Vyas and Priti P. Rege, "Malignancy Texture Classification in Digital Mammograms based on Chebyshev Moments and Logpolar Transformation", *ICGST-BIME Journal*, Volume 7, Issue 1, pp.29-35, December 2007.
- [23]. J.G. Zhang, T. Tan, "Brief Review of Invariant Texture Analysis Methods, in Pattern Recognition", Vol. 35, No.3, pp.735-747, 2002.
- [24]. Jan flusser, "moment invariants in image analysis, transactions on engineering, computing and technology", Vol.11, pp. 196-201, February, 2006.
- [25]. Jie Wei, "Shape Indexing and Recognition Based on Regional Analysis", *IEEE TRANSACTIONS ON MULTIMEDIA*, VOL. 9, NO. 5, pp.1-13, AUGUST 2007.
- [26]. <http://www.imaginis.com/breast-health/breast-cancer-statistics-on-incidence-survival-and-screening-2#breast-cancer-cases-deaths-per-year-u-s-and-world>.
- [27]. Breast Cancer in Men. A complete patient's guide. <http://www.breastdoctor.com/breast/men/cancer.htm>.
- [28]. Breast Cancer in Men. Male breast cancer information center. <http://interact.withus.com/interact/mbc/>.
- [29]. M. Karnan, K. Thangavel, "Automatic detection of the breast border and nipple position on digital mammograms using genetic algorithm for asymmetry approach to detection of microcalcifications". *Computer Methods and Programs in Biomedicine* Vol. 87, no.1, pp 12-20, 2007.
- [30]. T.O. Gulsrud, "Texture analysis of Digital Mammograms," Ph.D. Thesis, Aalborg University, Stavanger, USA, pp.30-32, 2000.
- [31]. D. C. He, L. Wang, "Texture classification using texture spectrum," *Pattern Recognition*, vol. 23, pp 905-910, 1990.
- [32]. M. H. Horng, "Texture feature coding method for texture classification," *Opt. Eng.*, Vol 42 (1), pp. 228-238, 2003.
- [33]. J.S. Suri, R.M. Rangayyan: *Recent Advances in Breast Imaging, Mammography, and Computer-Aided Diagnosis of Breast Cancer*, SPIE, Bellingham, Washington, USA, 2006.
- [34]. American College of Radiology (ACR): *ACR Breast Imaging Reporting and Data System*, 4th Edition, Reston, USA, 2003
- [35]. E.S. de Paredes: *Atlas of Mammography*, 3rd Edition, Lippincott Williams & Wilkins, Philadelphia, USA, 2007.
- [36]. I. H. Witten and E. Frank, *Data Mining, Practical Machine Learning Tools and Techniques*, 2nd ed. San Francisco: Morgan Kaufmann, 2005.
- [37]. M. Heath, "The Digital Database for Screening Mammography", *Proceedings of the 5th International Workshop on Digital Mammography*, Toronto, Canada, pp. 212-218, 2000.
- [38]. L.M. Bruce, R.R. Adhami: "Classifying Mammographic Mass Shapes Using the Wavelet Transform Modulus-Maxima Method", *IEEE Transactions on Medical Imaging*, Vol. 18, Issue 12, pp. 1170-1177, 1999.