An Efficient Hybrid Detection System for Abnormal Masses in Digital and Analog Mammogram

Ahmed S. Rashad[1], Rowayda A. Sadek[1,2] & Sherif Abdel Aziz El-Sherif[3]

[1] College of Engineering and Technology, Arab Academy for Science, Technology& Maritime Transport, Cairo, Egypt

[2] Faculty of Computers and Information, Helwan University, Cairo, Egypt

[3] Surgical Oncology Dept, National Cancer Institute, Cairo University, Cairo, Egypt

Correspondence: Ahmed S. Rashad, Dept. of Computer Engineering, College of Engineering and Technology, Arab Academy for Science, Technology& Maritime Transport, Sheraton Heliopolis, Cairo, Egypt. Tel: 20-111-492-1464. E-mail: eng.a.saaid@gmail.com

Abstract

Mammography is a type of radiography used on the breasts as screening method for women. The indicators for breast cancer aremasses and calcifications. Breast cancer screenings show that radiologists miss 8%–20% of the tumors. For this reason, development of systems for computer-aided detection (CAD) and computer-aided diagnosis (CADx) algorithms is the concern of a lot of researches currently being done. CAD and CADx algorithms assist radiologists in the decision between follow up and biopsy phases. An intelligent Image Processing Technique employed in systems that can help the radiology in detecting abnormal masses. This paper presents a general framework for mammography that will provide advantages for managing information and simplifying process in each layer for imaging technique. A method has been developed to make supporting tools used a framework as a reference model. This method will automatically segment and detect abnormal masses in analog and digital mammography images and compare between results.

Keywords: mammography, cancer, calcifications, cad, cadx

1. Introduction

Breast cancer ranks as the most common form of cancer and second-leading cause of cancer death among women (O.Whi-Vin et al., 2009). Detection of clustered microcalcifications in mammographic X-ray images in spite of absence of masses helps in diagnosis of early breast cancer. Those clustered microcalcifications are an indicator for early detection of breast cancer (Eddaoudi, Regragui, & Laraki, 2006). Groups of small calcifications huddled together, called "clusters of micro calcifications", are associated with extra breast cell activity (Jasjit & Rangaraj, 2006). Early detection can help reduce the number of deaths from breast cancer among women. A screening program based on mammography is considered one of the best and popular methodsbecause its fast and cost effective (Júlia, Deserno, & Araújo, 2008; Bentley, 2007). Mammography is the process of using low-energy-X-rays to examine the human breast and is used as a diagnostic and a screening tool. The goal of mammography is the early detection of breast cancer, typically through detection of characteristic masses and/or microcalcifications (Jasjit & Rangaraj, 2006).

Mammography uses different technologies such as analog mammography which uses x-ray to record images on film using an x-ray cassette. Films are then "developed" produced, put on a light box and then read by the Radiologist. Digital mammograms were first cleared by the FDA for use in 2003. The process is the same as the analog mammogram for the patient, except that it takes about half the time. The digital nature of the mammogram allows the image to be viewed on a monitor or printed in high resolution and can be enhanced (lightened or darkened) before they are printed on film (National Cancer Institute (NCI) Web site). Digital mammography offers the advantages of electronic film archiving, faster results, fewer patient callbacks and more shades of gray for analysis (The radiology information resource for patients). Determination of the existence of a benign or malignant disease cannot usually be done by initial mammographic or MRI images alone. If there is a suspicious lesion, the radiologist may recommend further diagnostic studies.

Interpretations of mammograms may be difficult because of the fact that; a normal breast is different for each woman (Maitra, Nag, & Bandyopadhyay, 2011). Also, presence of powder or salve on the breasts or previous breast surgery may compromise the appearance of the image.

High rates of false negative cases given by radiologists interpreting mammogram was proven by recent studies. (Highnam & Brady, 1999). Estimates indicate that between 10%-30% of breast radiologists miss cancers during routine screening (Baines, McFarlane, & Miller, 1990; Wallis, Walsh, & Lee, 1991). A variety of screening techniques have been developed to improve the accuracy of interpretation.

Radiologists can improve their performance; with the advances of digital image processing; with computer-aided diagnosis (CAD) system. Radiologists have the opportunity to use CAD system as an objective "second reader" to improve the sensitivity of mammography interpretation. On the average, computer-aided diagnosis (CAD) system can increase the reader's sensitivity by 10%. (Liyang, Yongyi, & Robert, 2009). The overall performance of any automated image analysis system is essentially affected by image segmentation. Imaging techniques play an important role on mammogram images, especially in abnormal areas that cannot be physically felt but can be seen or processed on an analog mammogram or with ultrasound (National Cancer Institute (NCI) Web site). Here by it appears the importance of developing a framework as a reference model for researchers to implement their techniques, without development overheads in computer-based diagnostic systems and improve drawbacks associated with this imaging technique.

Various works have been formulated to develop CAD and CADX. Many researchers used different approaches to do the classification of masses. The classification step is very important for computer aided diagnosis (CAD) system and its performance. Many researchers apply the detection techniques to detect the abnormal masses such as morphological operators and fuzzy c–means clustering (SahebBasha & Satya, 2009). Considering the textured nature of the mammographic images, we were interested in methods of gray levels reduction which do not degrade the quality of textures such as with rank coding method already used in satellite image processing (Hanifi et al., 2009) gray levels reduction method is more efficient and need less processing time.

This proposed paper is divided into four sections; the first one is devoted to the presentation of proposed generic framework. The second one is devoted to the presentation of proposed CAD. The third part shows the experimental results and discussions. Conclusion and future work are in the last section.

2. Proposed Framwork

Mammograms are medical images that are difficult to interpret, thus a preprocessing phase is needed in order to improve the image quality and make the segmentation results more accurate to help in detection and diagnosis phases. So this shows the importance of developing a generic mammography framework for CAD and CADX. This framework is proposed to generalize the development in the CAD and CADX by generalizing each module as a protocol layer. Each layer has a clear function to be adapted as well as their input and outputs. The main layers in the generic mammography framework are: image acquisition, segmentation "Global ROI", enhancement and abnormal mass detection "Local ROI" as it shown in Figure 1.

The first layer is image acquisition which includes digital and analogue methods (National Cancer Institute (NCI) Web site). The inputs are analogue images or digital images by different formats and methods. The Outputs are uniformed image that take in consideration standard format and size. The second layer is segmentation "Global ROI" which includes two sub layers. The first sublayer is concerned with removal the noise and labels from background "artifact suppression, background separation". The second sublayer is concerned with removal of pectoralis muscle region which is very important in analogue images because detection algorithms consider it as a mass. This sublayer is optional for digital images. The third layer "Enhancement". This layer plays an important role in generic mammography framework by improving the quality of images which make the detection process more efficient. The fourth layer is abnormal mass detection "Local ROI" this layer is divided into two sublayer. The first sublayer is concerned with small mass detection such as microclasifications and clustered microcalsificationswhich are indicators for breast cancers. The second sublayer is concerned with large mass detection that may be malignant or benign.
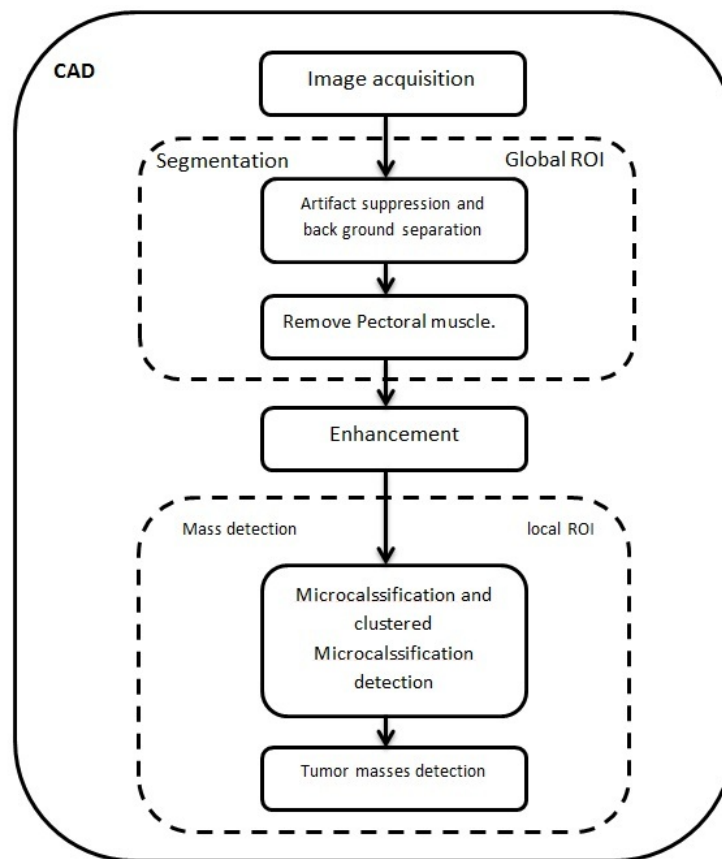
Figure 1. The proposed generic framework for mammography

3. The Proposed CAD

The paper proposed an efficient CAD to accurately detect small and large masses for both digital and analogue dataset. The CAD takes generic mammography framework as a reference model in development cycle. The next section will discuss each layer in details and how the implementation was done.

3.1 Layer 1: Image Acquisition

The databases used in this layer were produced from analogue and digital mammography.

3.2 Layer 2: Mammogram Segmentation (Artifact Suppression, Back Ground Separation and Pectoral Muscle Region Removal)

The first step involves the removal of unwanted areas in the background of the mammography images and pectoralis muscle. The main objective of this process is to improve the quality of the image, to make it ready for further processing and make the segmentation results more accurate. The step is divided into two parts: Artifact suppression background separation and Removal of Pectoral muscle.

3.2.1 Mammogram Segmentation (Artifact Suppression, Background Separation)

Radioopaque artifacts such as wedges and labels in the mammogram images are removed using threshold and morphological operations. Transform the grayscale images into binary format. After that some morphological operations such as dilation, erosion, opening and closing are performed on the binary images. The resulting binary image is multiplied with the original mammogram image to form the final grayscale image. This algorithm is used for suppression of artifacts, labels and wedges (JawadNagi et al., 2010). This method will be used as first stage in preprocessing and will show its impact in case study section.

3.2.2 Pectoralis Muscle Region Removal

The reliability of boundary matching may be increased by extracting the pectoralis muscle from the breast region. The pectoralis muscle appears as a bright triangular region in the image corner towards the chest wall and the bottom of the breast region. A histogram-based threshold technique is used to separate the pectoralis muscle region (Kwok et al., 2004). This method will be used as second stage in preprocessing and its impact will be examined in case study section.

3.3 Layer 3: Enhancement

The primary goal of mammography screening is early detection of small, non-palpable masses. Pathological changes of the breast being variable and poorly visible in low contrast and noisy mammograms make interpretation of the mammograms difficult. Several algorithms for image enhancement were proposed in the literature to increase the visibility of these features. Image enhancement is an important step in Computer-Aided Detection (CAD) systems for automated analysis of mammograms. Radiologists make the final decision as these systems work only as a second reader. When using CAD systems as an aid, it was found that radiologists' accuracy for detecting breast cancer is improved, as shown by some studies. Image enhancement techniques are used to improve an image, where "improve" is sometimes defined objectively (e.g., increase the signal-to-noise ratio), and sometimes subjectively (e.g., make certain features easier to see by modifying the colors or intensities). Contrast enhancement and equalization of histogram are essential technical aids in applications where human visual perception remains the primary approach to extract relevant information from images. Which increase the detection ability of mass signs at earlier stages as compared with the original and unprocessed digitized mammograms (Nagraja et al., 2011).

3.4 Layer 4: Abnormal Mass Detection

Mammography is at present one of the available methods for early detection of abnormalities related to breast cancer. The indicators for breast cancer are masses and calcifications. Masses are defined by their shapes and margin properties. According to the shape and boundary characteristics of masses, it can be further sub divided into speculated masses (SPIC), circumscribed masses (CIRC), and other masses (MISC). Calcifications are calcium deposits within breast tissue. They appear as white spots or flecks on a mammogram and are usually so small that you can't feel them (Guliato et al., 2008). The size of a mammographic mass can vary in a wide range (~3 mm – ~50 mm). The intensity based mass detector defines two size classes ("small" and "large") and uses different methods for each. Most real-life mammographic masses belong to the "small" class so the small mass detector is the critical part of the system. Masses belonging to the "large" class are rare and usually easy to detect. Therefore the large mass detector is a less-frequently-used, simpler, but necessary part (Tóth, Takács, & Pataki, 2008).

3.4.1 Detection of Large Masses

The identification technique is divided into two distinct parts i.e. Formation of Homogeneous Blocks and Color Quantization after preprocessing (Maitra, Nag, & Bandyopadhyay, 2011).

3.4.1.1 Formation of Homogeneous Blocks

By analyzing mammogram image, the mammogramis segmented into very small blocks. To reduce the complexity of the algorithm, first the image degenerates into 2X2 pixel blocks. Intensity or the pixel value of the blocks is checked to obtain the pixel value which presents maximum occurrence within the block. The system propagates the value in the adjacent pixels of the block. Now the entire block contains the same pixel value. So, the whole mammogram image now consists of 2X2 homogeneous blocks. The 2X2 homogeneous blocks are arranged to obtain a 4X4 pixel block containing 4 arranged 2x2 blocks. A similar process is continued with this 4X4 block as done with the initial 2X2 blocks. Now, the whole mammogram image consists of 4X4 homogeneous. The same process is repeated to produce mammogram image of 8X8 homogeneous blocks by arranging 4 individual 4X4 blocks likewise as stated above.

3.4.1.2 Color Quantization

Uniform color quantization technique is used to break the color space of mammogram image into eight equal sized regions. Although color quantization is a lossy process but it is used. The idea is sampling the original mammogram image for color statistics, choosing a color map based on those statistics and mapping the colors to their representative properties in the color map. The 8X8 homogeneous mammogram image is now clearly segmented into different color regions and each region representing specific part and properties.

3.4.2 Detection of Small Masses (Microcalssifications)

Several techniques can be adopted to detect microcalcifications. Texture-analysis is one of the efficient tools for microcalcifications detection. The most used texture analysis method for textural features extracted is the co-occurrence matrix (Eddaoudi & Regragui, 2011). The major weakness of the analysis of textures by the co-occurrence matrix is the computing time. The goal of this new method is to highlight textured information required in the mammograms by coding its textures and removing superfluous information in order to decrease the quantity of data to analyze and thus decrease the computation time (Haralick, 1979). The aim is to show that textual information is stronger in the coded image than in the original image. This coding consists, in assigning to each pixel of the image, a value of code and not a gray level. Indeed, the original image is transformed to create an intermediate image which is the coded one from which the texture characteristics can be extracted through the matrix.

4. Experimental Results and Discussions

The CAD in this paper is validated by using a benchmark database and advanced local database. The following is a list of some of the databases that are commonly used: The methodology presented in this work was applied on the complete miniMIAS database (Suckling et al., 1994).

The local dataset contains mammograms obtained from National Cancer Institute Cairo University consist of 26 patient which support 45 case. The two standard views (medio-lateral oblique and cranio-caudal) for each case. The images of the database originate from a digital mammography.

The proposed CAD is applied on 45 random case from mini MIAS database containing 12 normal cases, 13 microcalcification and 20 abnormal masses, and 45 case from local dataset contains 12 normal cases, 13 microcalcification and 20 abnormal masses. The following experimental studies were carried out on both databases (benchmark and local ones). The experiments considered different cases such as normal breast, breast containing microcalssifications or abnormal small and large masses.

The first phase was carried out using the benchmark analog database Figure 2. Figure 2(a) shows original image with noises and labels, and Figure 2(b) shows background separation from any unwanted signs. Figure 2(c) shows the output of muscle removal that may be detected as mass in mass detection layer, while Figure 2(d) shows the output of enhancement methods that was applied on image to improve its quality and make detection layer more efficient and Figure 2(c) shows the output of detection algorithms for both small and large masses.

a   b   c   d   e

(i) Normal breast

a   b   c   d   e

(ii) Abnormal large masses

a   b   c   d   e
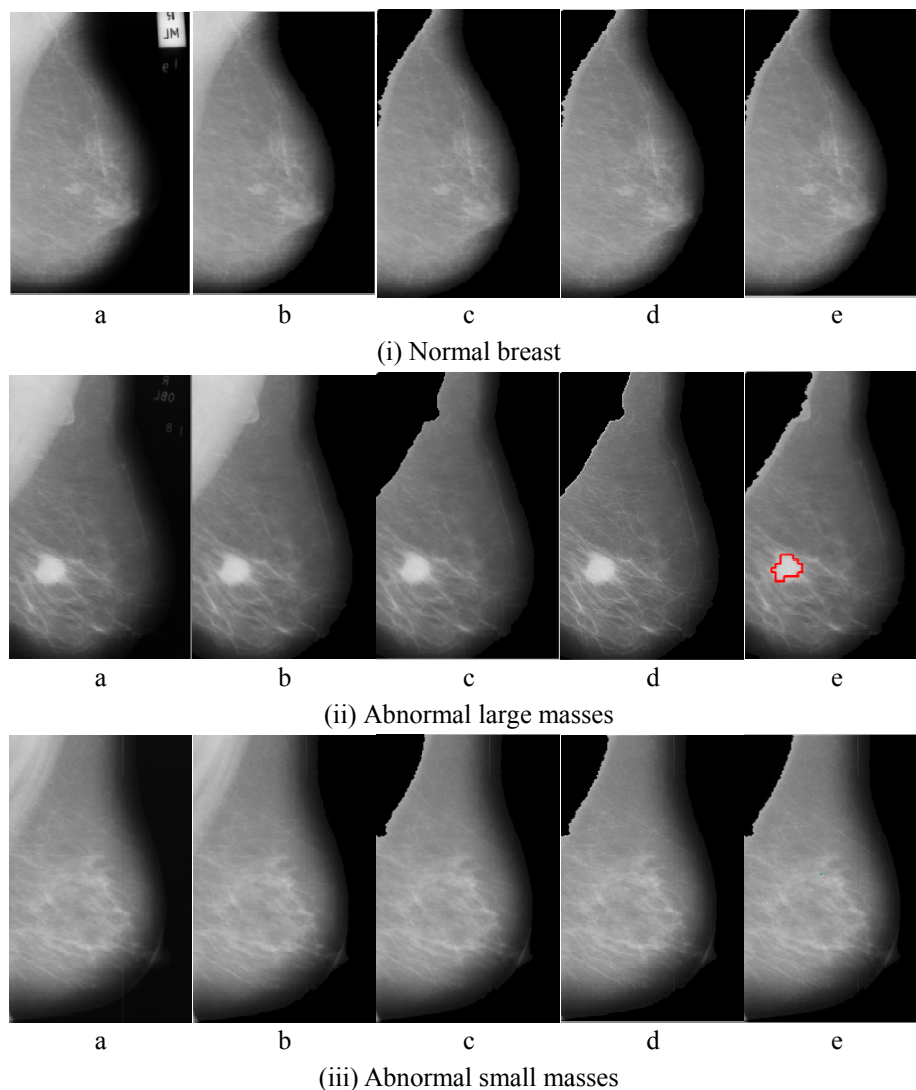
(iii) Abnormal small masses

Figure 2. For analog dataset: (a) original image; (b) global ROI; (c) removed muscle; (d) enhanced image; (e) local ROI

The second phase is carried out using the local database. Since the local dataset is a new DB; there was a need to benchmark it. Subjected by a radiology expert to classify the image; which is normal, has small masses or has large masses. The local dataset is produced from digital mammogram whit high accuracy. So the removal of muscle may notbe needed in this case (optional) because it may help to detect weather the enlarged lymph nodes were diseased or not. Figure 3 shows the digital dataset cases. Figure 3(a) shows original image with labels, while Figure 2(b) shows background separation from any unwanted labels. The output of enhancement methods is the output of detection algorithms for both small and large masses.

(a)   (b)

(i) Normal breast

(a)   (b)

(ii) Abnormal large masses
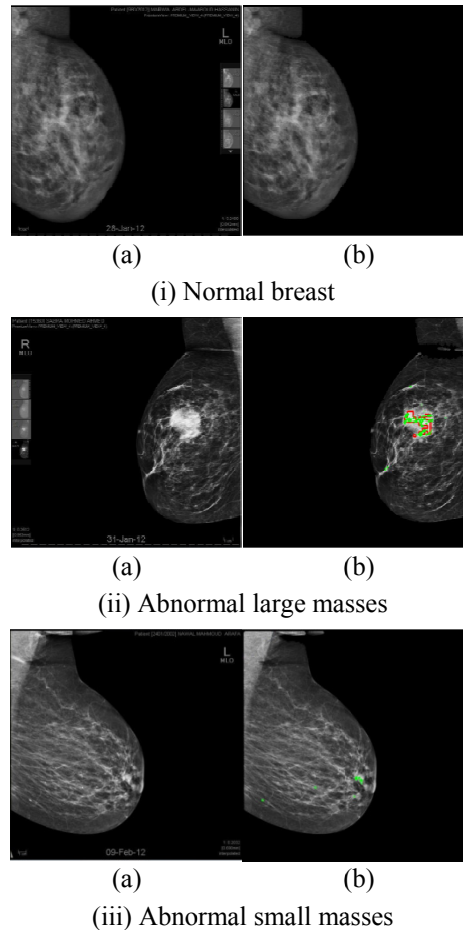

(a)   (b)

(iii) Abnormal small masses

Figure 3. For Digital Dataset: (a) Original Image (b) op of abnormal mass detection algorithm

A receiver operating characteristic (ROC or simply ROC curve, is a plot of the sensitivity, or true positive rate (TPR), vs. false positive rate (FPR) for a binary classifier system. From TPR and FPR the accuracy can be calculated.

$$True\ positive\ rate\ (sensitivity) = TP / P \tag{1}$$

Where TP is true positive value and P is positive value.

$$False\ positive\ rate\ (specificity) = FP / N \tag{2}$$

Where FP is false positive value and N is negative value.

$$Accuracy = (TP+TN)/ (P+N) \tag{3}$$

ROC analysis has been used in medicine, radiology, biometrics, and other areas for many decades and is increasingly used in machine learning and data mining researches (Zweig & Campbell, 1993), when apply ROC equations on analogue dataset (45 case). That was found 24 cases are true positive from 33 cases by substitution in Equation (1) the result was 73.33% and 3 cases are false positive from 12 cases by substitution in Equation (2) the result was 25%. From the previous results the accuracy will equal 73.3% by substitution in Equation (3).

When apply ROC equations on digital dataset (45 cases). That was found 26 cases are true positive from 33 cases by substitution in Equation (1) the result was 78.78% and 1 cases are false positive from 12 cases by substitution in Equation (2) the result was 8.33%, from the previous results the accuracy will equal 82.22% by substitution in Equation (3). The experimental results showed that the detection algorithm which applied on digital images give more accurate results than analog images. This mean the digital images are more efficient in our CAD which gives more accurate results that are help the radiology.

5. Conclusion and Future Work

There are two key contributions of the work presented in this paper. The first of these is fully automated detection algorithm. The second contribute is algorithms performance evaluation on both digital and along images.

Due to multiple features used for mass detection, some abnormalities may be missed or misinterpreted. There is also a number of false positive findings and therefore a lot of unnecessary biopsies. Computer-aided detection algorithms have been developed to help radiologists to give an accurate diagnosis and to reduce the number of false positives. In this paper, algorithms that are commonly used which applied on different dataset show that the result of digital mammography is better than analog. The output of ROC equation accuracy is rarely above 73.33 % for analog dataset and 82.22 % for digital dataset. The CAD system should use digital mammogram which gives more accurate result than analog mammogram as shown on ROC results.

In future work, currently popular Artificial Intelligent techniques could be used into the segmentation and classification process to increase the efficiency and accuracy of computer-aided detection tools. These techniques such as neural network and support vector machine could also be applied to detection and classifications of breast tumor in mammograms and other modalities.

Acknowledgements

References

Baines, C. J., McFarlane, D. V., & Miller, A. B. (1990). The Role Of The Reference Radiologist: Estimates Of Interobserver Agreement And Potential Delay In Cancer Detection In The National Screening Study. *Invest Radiol, 25*, 971-976. http://dx.doi.org/10.1097/00004424-199009000-00002

Bentley, K. (2007). Mammography image quality: Analysis of evaluation criteria using pectoral muscle presentation, Radiography.

Breast Cancer Facts and Figures. (2009-2010). American Cancer Society, Inc.

Eddaoudi, F., & Regragui, F. (2011). Microcalcifications Detection in Mammographic Images Using TextureCoding. *Applied Mathematical Sciences, 5*(8), 381-393.

Eddaoudi, F., Regragui, F., Laraki, K. (2006). Characterization of normal mammograms based on the Statistical Features. Proc. Second International Symposium on Communications, Control and Signal Processing (ISCCSP), Marrakech.

Guliato, D., Rangayyan, R. M., Carvalho, J. D., & Santiago, S. A. (2008). Polygonal Modeling of Contours of Breast Tumors with the Preservation of Spicules. *IEEE Trans. Biomed. Eng., 55*(1), 14-20. http://dx.doi.org/10.1109/TBME.2007.899310

Hanifi, M., Sedes, F., Aboutajdine, D., & Lasfar, D. (2009). A New Approach for Coding Satellite Images: Rank Coding. *International Journal of Computational Science, 3*(2), 151-163.

Haralick, R. M. (1979). Statistical and structural approaches to texture. *Proceedings of the IEEE, 67*(5), 786-804. http://dx.doi.org/10.1109/PROC.1979.11328

Jasjit, S. S., & Rangaraj, M. R. (2006). *Recent Advances in Breast Imaging, Mammography and Computer-Aided Diagnosis of Breast Cancer*. Belligham, Washington, Spie Press.

Júlia, E. E. O., Deserno, T. M., & Araújo, A. (2008). Breast Lesions Classification applied to a reference database. 2nd International Conference: E-Medical Systems, Tunisia.

Kwok, S. M., Chandrasekhar, R., Attikiouzel, Y., & Rickard, M. T. (2004). Automatic Pectoral Muscle Segmentation on Mediolateral Oblique View Mammograms. *IEEE Transactions on Medical Imaging, 23*(9), 232-246. http://dx.doi.org/10.1109/TMI.2004.830529

Liyang, W., Yongyi, Y., & Robert, M. N. (2009). Microcalcification Classification Assisted by Content-Based Image Retrieval for Breast Cancer Diagnosis. *Pattern Recognition, 42*, 1126-1132. http://dx.doi.org/10.1016/j.patcog.2008.08.028

Maitra, I. K., Nag, S., & Bandyopadhyay, S. K. (2011). Identification of Abnormal Masses in Digital Mammography Images. *International Journal of Computer Graphics, 2*(1), 17-29.

Nagi, K., Sameem, A. K., Nagi, F., & Ahmed, S. K. (2010). Automated Breast Profile Segmentation for ROI Detection Using Digital Mammograms. IEEE EMBS Conference on Biomedical Engineering & Sciences (IECBES 2010), Kuala Lumpur, Malaysia, November 2010.

Nagraja, M., Rao, S., Bhattacharya, M., & Bommanalli, S. (2011). Enhancing Image Contrast Of Mammogram & Equalization of Histograms. International Journal of Engineering Science and Technology (Ijest), 3 No.

National Cancer Institute (NCI). Retrieved from http://www.cancernet.gov

Saheb B. S., & Satya, P. K. (2009). Automatic Detection of Breast Cancer Mass in Mammograms using Morphological Operators and Fuzzy C – Means Clustering. *Journal of Theoretical and Applied Information Technology,* 704-709.

Suckling, J., Parker, J., Dance, D., Astley, S., Hutt, I., Boggis, C., …, Kok, S. (1994). The mammographic image analysis society digital mammogram database. In: ExerptaMedica. International Congress Series, pp. 375-378.

The radiology information resource for patients. (n.d.) Retrieved from RadiologyInfo.org

Tóth, N., Takács, G., & Pataki, B. (2008). Mass Detection in Mammograms Combining Two Methods. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.78.8512

Wallis, M. G., Walsh, M. T., & Lee, J. R. (1991). A Review Of False Negative Mammography In A Symptomatic Population. *ClinRadiol, 44*, 13-15. http://dx.doi.org/10.1016/S0009-9260(05)80218-1

Whi-Vin, O., KwangGi, K., Young-Jae, K., HanSung, K., JungSil, R., & WooKyung, M. (2009). Detection of Microcalcifications in Digital Mammograms Using Foveal Method. *Journal of Korean Society of Medical Informatics, 15*(1), 165-172.

Zweig, M. H., & Campbell, G. (1993). Receiver Operating Characteristic (ROC) Plots: A Fundamental Evaluation Tool In Clinical Medicine. *Clin. Chem., 39*(4), 56-577.