Original Research Article

Biotechnology

International Journal of Pharma and Bio Sciences

ISSN
0975-6299

EARLY STAGE DETECTION OF BREAST CANCER USING FEATURE EXTRACTION TECHNIQUES

SHAIK NASEERA[1], G.K. RAJINI[2], ARIVUSELVAN. K[3] AND T. SREENIVASULU REDDY[4]

[1]School of Computer Engineering , VIT University, Vellore, India.
[2]School of Electrical Engineering, VIT University, Vellore, India.
[3]School of Information Technology, VIT University, Vellore, India.
[4]Sri Venkateswara University, Tirupati, India.

ABSTRACT

This paper presents an effective and efficient approach for early detection of breast cancer in digital mammogram images, in which masses and microcalcifications appear in the form of clusters with high intensity compared to their neighborhood pixels. The presence of these clusters is considered as a prominant indicator of malignant and benign types of breast cancer and the detection of this is used to treat/prevent the disease at an early stage. In order to implement this, we propose a new algorithm that uses image enhancement by wavelet transform and adaptive histogram equalization technique followed by Segmentation with Border extraction. We compared the proposed algorithm with algorithms that used conventional watershed segmentation for detecting the masses and microcalcification efficiently.

KEYWORDS: **Breast Cancer, Mammogram , Wavelet Transform , Segmentation, Malignant and Benign, microcalcification**

SHAIK NASEERA
School of Computing Science and Engineering,
VIT University, Vellore, India.

***Corresponding author**

I. INTRODUCTION

Breast cancer is the most alarming type of disease confront by women around the globe. Early detection and diagnosis of breast cancer is a challenging issue which increases the survival rate and chances of thorough recovery. Some of the abnormal features is detected by several image processing techniques are Masses, Microcalcifications, Bilateral Asymmetry and Architectural distortion. These generally indicated the malignant and benign type of cancer. Often Benign masses are well-defined sharp border, while malignant masses possess ill-defined, spiculated or microlobulated borders. Calcifications are deposits of calcium in breast tissue. Benign calcifications are usually larger and coarser with round and smooth contours. Malignant calcifications tend to be numerous, clustered, small, varying in size and shape, angular, irregularly shaped and branching in orientation. Microcalcification corresponds to high frequency image spectrum and is detected by Wavelet Transform (WT) an emerging technique to solve the inability of Fourier Transform[1] . WT is a mathematical tool which decomposes image into four sub-bands and allows good resolution in time and frequency components. The four subbands are approximation(LL), horizontal(LH), vertical(HL) and detail coefficients(HH).Low frequency sub-bands are eliminated and image is reconstructed from high frequency bands as the masses and microcalcification corresponds to high frequency components. Pyramid segmentation and decision tree classification were used to detect the masses[2] . Bilateral asymmetry was detected using B-spline interpolation[3] . Bilateral breast volume asymmetry in screening mammogram is calculated as a potential marker of breast cancer analysis[4] . Breast shape is assumed as a conical shape for volume calculations[5] . In this paper, we are applying image processing techniques such as feature extraction, segmentation, wavelet transformation techniques for mammogram images. The proposed new algorithm uses image enhancement by wavelet transform and adaptive histogram equalization technique followed by Segmentation with Border extraction. This paper is organized as follows: Section II provides related work on the topic of interest. Section III describes implementation techniques using MATLAB AND LABVIEW, Section IV presents results and Section V concludes the superiority of the algorithm.

II. RELATED WORK

The most common cause of death in women is Breast cancer and also the second leading cause of cancer deaths in the entire world (after lung cancer)[9] . On an average 1,82,000 new cases of breast cancer are diagnosed and 46,000 women die due to breast cancer each year in the US[10] . Early detection is the crucial step in breast cancer diagnosis and treatment, because there is no effective way to prevent or treatment for breast cancer till now. Now days, X-ray mammography is the most viable technique followed in clinical laboratories and practice for its affordable cost. Computer Aided Diagnosis (CAD) is being used in the screening process to support radiologists and internists for diagnosis to improve the accuracy and efficiency of mammogram examination. CAD systems are generally used to support the interpretation of medical images. CAD schemes such as computer aided detection (CADe) and computer-aided diagnosis (CADx) are most popular in this. CADe is used for the identification of the location of suspect regions and CADx is targeted for characterization (i.e., malignancy versus benignity)[9] . R. Guzmán-Cabrera et. al. presented an approach to analyze digital mammograms more effectively using texture segmentation for the detection of early stage tumours[11]. The proposed algorithm was tested over several images taken from the digital database for screening mammography for cancer research and diagnosis. This technique is found to be more suitable for distinguishing masses and microcalcification from the background tissue using morphological operators and then extract them through machine learning techniques and a clustering algorithm for intensity-based segmentation. Vijayalakshmi et.al has done[13] cytological grading on aspirates of breast carcinomas and histological grading of corresponding specimens using Scarff Bloom Richardson's grading and the lymph nodes were also studied for detection of metastasis. Nagaraju et. al applied an innovative method consisting of three steps such as normalizing the regions in the breast images through uniform distribution of histogram equalization, Application of fuzzy logic to remove ambiguity in the misclassification region and finally a new Weight is applied to the extended OTSU method for early stage detection of breast cancer[14].

III. IMPLEMENTATION TECHNIQUES USING MATLAB AND LABVIEW

A. Image Enhancement

Image enhancement is defined as a process of improving the image quality to a better and more understandable level for feature extraction or image interpretation which is more suitable for a particular application[12]. These techniques include intensity and contrast manipulation, noise reduction, background removal, histogram conversion, edge sharpening, filtering etc. Gaussian pyramid is the process of filtering and resampling technique which is applied for mammogram images. The standard resolution of image is 1914×2294. The hierarchy consists of two levels (0-1).

Figure 1
Enhancement Technique

A. Algorithm
Step 1: Gaussian filter is applied to the mammogram image for removal of noise. It is then scaled down by cubic spline resampling technique using labview to bring it to level-1 image i.e., half width and height. The process of transforming a sampled image from one co-ordinate system to another is called image resampling. The mapping function of the spatial transformation relates the two co-ordinate systems to each other[6]. Step 2: Gaussian filter is applied to level1 in labview and this image is up-scaled by using cubic spline interpolation technique. Interpolation is the process of digital zooming, is to increase the image pixel number. So, a low resolution image can be converted into a high resolution image which can be of high quality with good visual effect. Step 3: Wavelet fusion is applied using Daubechies wavelet using technical computing language called Matlab. The process of combining of two images into a single image that has the maximum information content than the input images is called image fusion. Fusion has been done by means of Daubechies which uses overlapping windows, so the high frequency spectrum reflects all the high frequency changes. It is essentially used because, masses and microcalcification corresponds to only high frequency components. Step 4: Adaptive Histogram Equalization (AHE) is applied to the fused image. It improves the brightness of the image. It is appropriate to adjust the local contrast and to fetch the clear details. Table 1. Shows the enhancement images with AHE and wavelet images of normal, benign and cancer stages

C. Segmentation
Segmentation is the process of partitioning an image into region that corresponds to identification of objects or relevant information which possess common property called intensity, thresholding etc. Good shape connectivity and matching can be easily obtained through segmentation techniques.

C.1. Region growing segmentation Algorithm

Step 1
Segmentation of masses is implemented by region growing technique. It is a simplest method that correctly separates the similar properties which we define. It is an iterative process that checks initial seed points neighboring pixels in order to determine whether the neighboring pixels can be added to the seed points or not. Threshold and intensity values are found out by histogram.

Step 2
Morphological operations for binary images provide a basic techniques and those operations for gray scale images require more sophisticated mathematical concepts for extracting image components. These

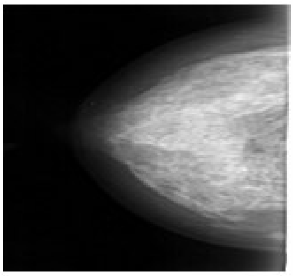
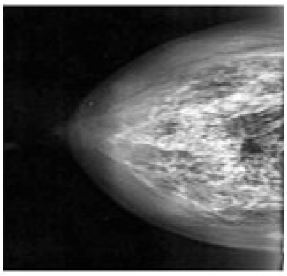
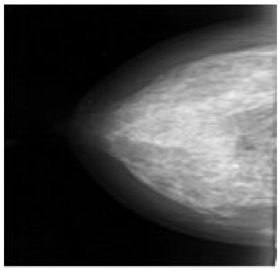
components are useful in the description and representation of region shape such as skeleton, boundaries etc. Dilation and erosion are the basic morphological processing operations. Both dilation and erosion are produced by the interaction of a structuring element. The structuring elements are a small set of sub images used to probe an analyzed image for properties of interest. Structuring element has both shape and an origin. Dilation enhances objects in a binary image while erosion shrinks which is controlled by the structuring element. Gray scale dilation and erosion using 5×5 kernel as structuring element is implemented in Labview. Gray morphological operations are used to extract the edge of the image to enhanced noise immunity, improve the detection of image accuracy and to effectively identify the target parts in the image.

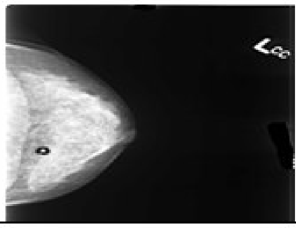
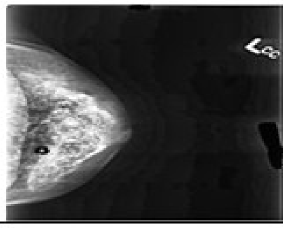
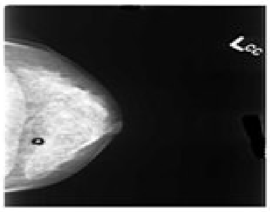
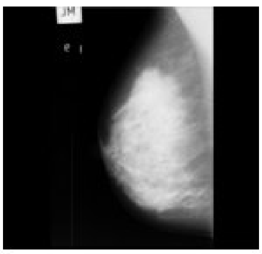
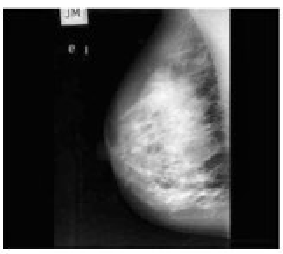
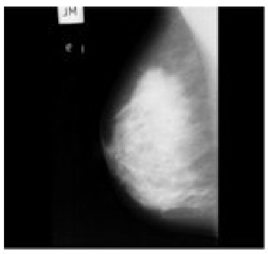
Step 3

Border extraction is usually done by canny edge detection technique. Edge detection is used to analyze objects in an image and for recognition. When the threshold level exceeds certain limit, the information of an image will get lost. But canny is the best algorithm as it is based on gradients and non-maxima suppression which filters noise, maintains integrity of the valid information and ensures greater positioning accuracy of the image. Obtain the margin of the mass, area, mean, standard deviation using magic wand tool in Labview.

Table 1

Histogram of Adaptive Histogram Equalization

	Original mammogram	AHE	Wavelet
Normal Stage			

	Original mammogram	Masses	Border extraction
Benign Stage			
Cancer Stage			

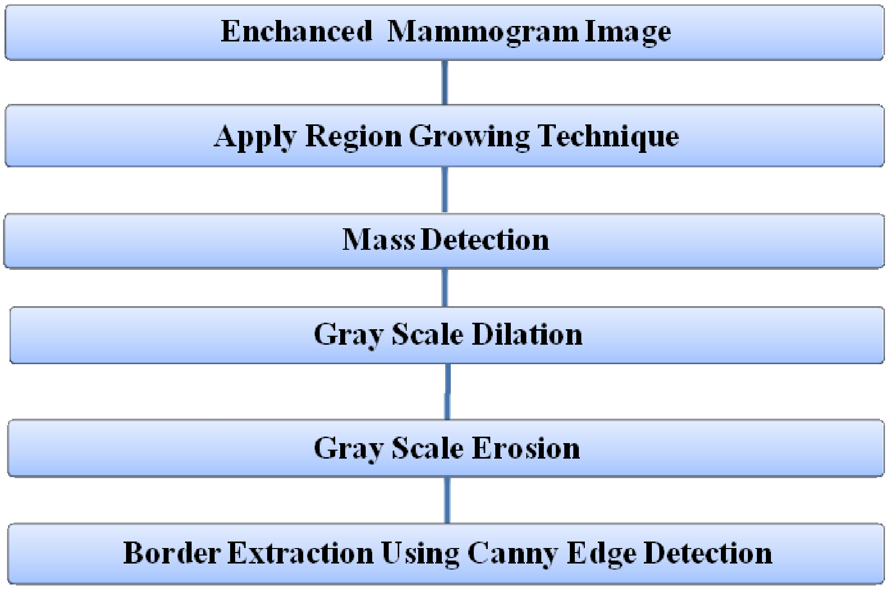
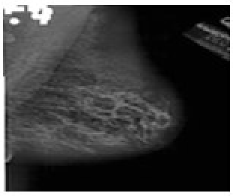
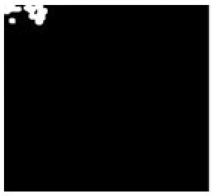
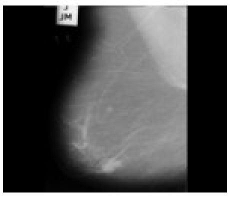
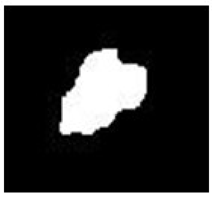
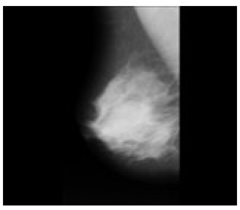
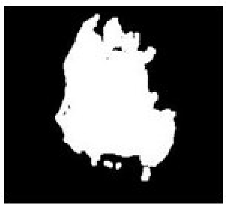
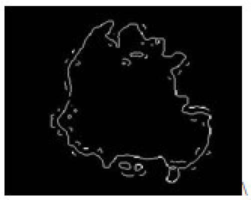


Figure 2
Flowchart for Region Growing Segmentation Technique

Table 2
Output images of region growing segmentation with masses and border extraction

Normal stage			No border
Benign stage			
Cancer stage			

C.2. Watershed Segmentation

In watershed segmentation, image enhancement techniques are used to avoid under-segmentation and noise removal technique is used to avoid over-segmentation. Watershed segmentation only allows binary image and these are called pre-processing techniques applied to the mammogram images. Using threshold segmentation, binary image can be obtained[7].

Algorithm

Step 1: Mammogram as an input image.
Step 2: Apply high pass filter for noise removal and median filter to enhance the quality of image
Step 3: Apply threshold and watershed segmentation
Step 4: Apply morphological operation

Step 5
Final output will be a mass region

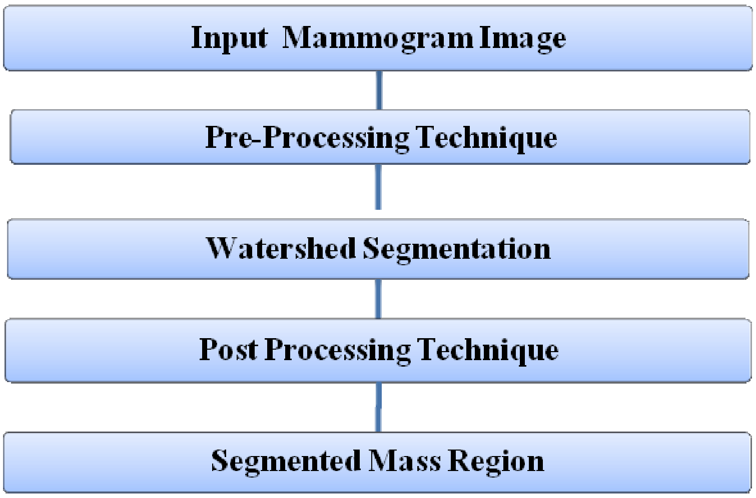
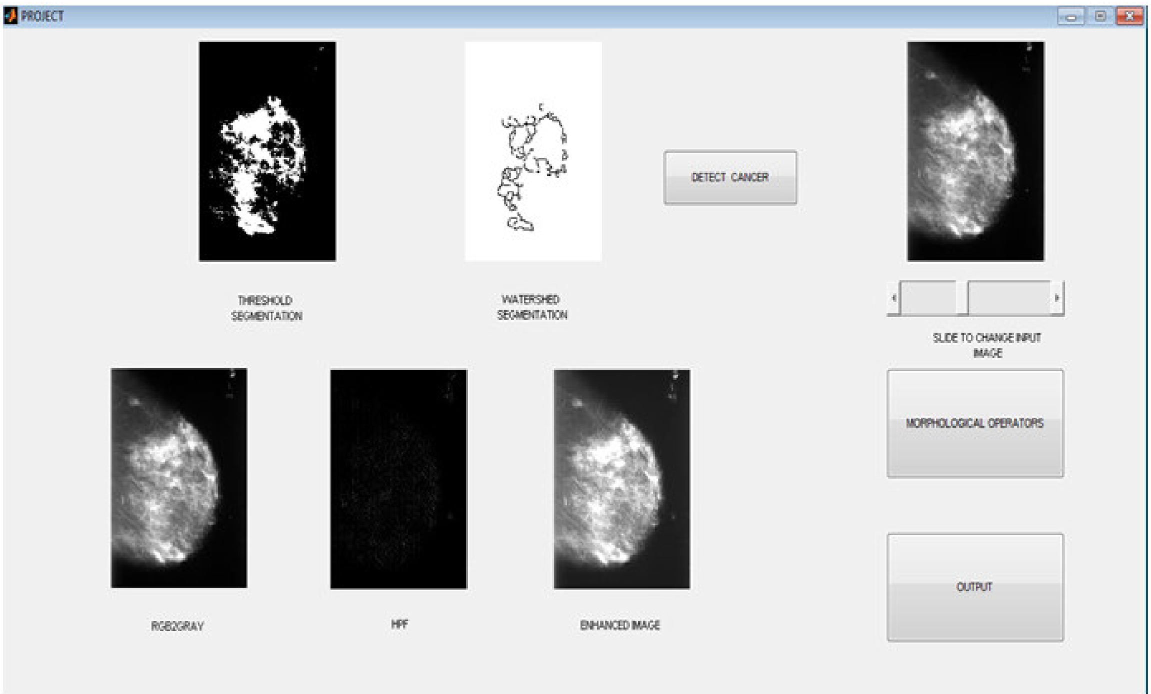


Figure 3
Flowchart for Watershed Segmentation Technique

Table 3
Output images of watershed segmentation with masses and segmented image.



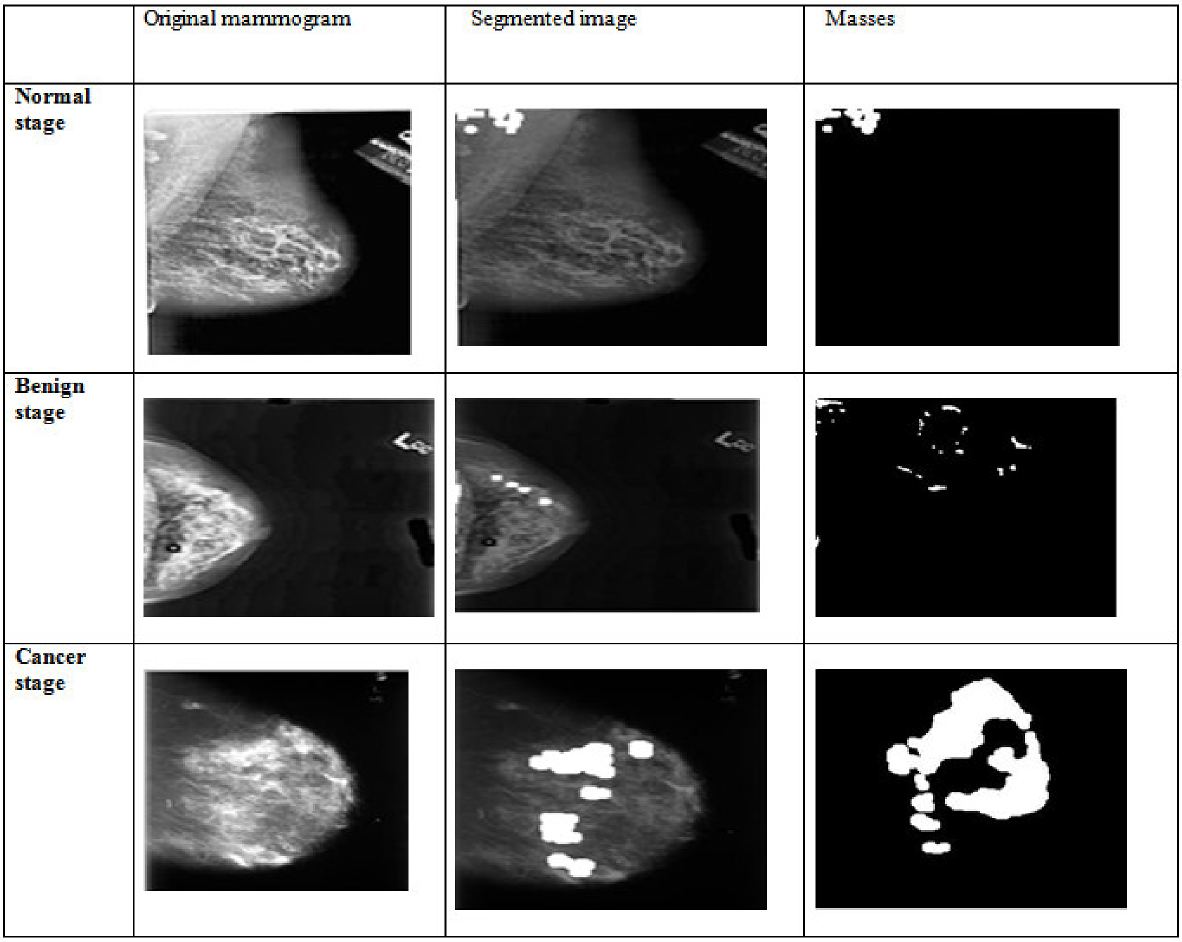


Figure 4
Implementation of Watershed Segmentation through MATLAB

IV. RESULTS

Table 4 shows the mean and standard deviation of speculated mass of processed mammogram images taken from DDSM and Mini-MIAS Database. The raw mammogram images are filtered using Gaussian filter. Later, wavelet fusion technique was applied because wavelet transformation extracts the high frequency coefficients of the image from which the detection of local structures at different scales is done. Denoised images obtained are subjected to Adaptive Histogram Equalization that limits the appearance of artifacts and noise. Image Segmentation based on Watershed algorithm and morphological operators gives the accurate means of earlier detection and classification of cancer stages. Finally border extraction of masses are done using Canny edge detection. Table 4 provides the state of masses present based on the Mean, Area and Standard deviation of Speculated, Circumscribed Microlobulated and ill defined mass present in mammogram images by using the proposed algorithm. The results shown in Table 4 demonstrates that better classification is realized for benign and malignant tissues.

Table 4
State of masses present based on the Mean, Area and Standard deviation

Mass Margin	Area (%)	Mean	Standard deviation	Mass state
Speculated	65.461	253.349	3.71953	Malignant
Circumscribed	8.349	252.27	4.60011	Benign
Microlobulated	35.86	251.073	5.1714	Malignant
Ill-defined	59.73	253.625	3.5074	Malignant

V. CONCLUSION

The proposed work presents the processing of number of patients images taken from DDSM and Mini-MIAS Database and we can conclude from the processed mammogram images is that standard deviation of speculated mass is less compared to the circumscribed mass. The proposed methodology helps the physician to identify the cancer at an early stage. We can identify from the border which indicates that benign mass (normal) has well define border while Malignant mass(cancer) has ill-defined or irregular shape or border which facilitates the physicians to further investigate the cases for early detection of breast cancer. In future work we would like to further study and classify more features of masses to identify the benign or malignant patients for early detection of breast cancer and complete recovery.

REFERENCES

1. Ted Wang C., Nicoloas Karayiannis B. Detection of microcalcifications in Digital Mammograms Using wavelets. IEEE Int. Trans. In Medical Imaging, 98-509,(1998).
2. Hewang Lin-Lin Huang., Xiao Jie Zhao., Automated detection of masses in Digital mammogram Based on Pyramid. IEEE Int. Conf in Wavelet analysis and pattern recognition. 183-187,(2007).
3. Jelena Bozek., Emil Dumic., Mislav Grgic. Bilateral Asymmetry Detection in Digital mammography using B-spline Interpolation. IEEE Int. Conf. in systems, signal and Image processing, 1-4,(2009).
4. Nevine Eltonsy H., Adel Elmaghraby S., Georgia Tourassi D. Bilateral Breast volume Asymmetry in screening Mammograms as a potential Marker of Breast cancer. Preliminary Experience. IEEE Int. Conf on Image Processing, 5-7,(2007).
5. Katariya R. N., Forrest A. P. M., Gravelle I. H. Breast volumes in cancer of the breast. British journal of cancer, 29(3), 270, (1974).
6. George Wolberg. Digital Image Warping. IEEE Computer Society Press, Los Alomitos, CA, USA, (1994).
7. Patil R. C., Bhalchandra A. S. Brain tumor extraction from MRI images using MATLAB. International Journal of Electronics, Communication & Soft Computing Science and Engineering, 2(1), 1-4, (2011).
8. Spandana P., Rao K. M., Rao B. V. V. S., Jwalasrikala J. Novel image processing techniques for early detection of breast cancer, mat lab and lab view implementation. In Point-of-Care Healthcare Technologies (PHT), IEEE, 105-108, (2013).
9. Deserno T. M. Fundamentals of biomedical image processing. In Biomedical Image Processing, Springer Berlin Heidelberg.1-51, (2011).
10. Cheng H. D., Cai X., Chen X., Hu L., Lou, X. Computer-aided detection and classification of microcalcifications in mammograms: a survey. Pattern recognition, 36(12), 2967-2991, (2003).
11. Guzmán-Cabrera R., Guzmán-Sepúlveda J. R., Torres-Cisneros M., May-Arrioja D. A., Ruiz-Pinales J., Ibarra-Manzano O. G., Parada, A. G. Digital image processing technique for breast cancer detection. International Journal of Thermophysics, 34(8-9), 1519-1531,(2013).
12. Gonzalez R. C, Woods R. E, and Eddins S. L. Digital image processing using MATLAB. Pearson Education India, (2004).
13. Vijayalaxmi S.P., Potekar R.M., Yelikar B.R., Karigoudar M.H. Comparative Study of Fine Needle Aspiration Cytology and Histopathology in Grading Breast Carcinoma. International Journal of Pharma and Bio Sciences, 5(2):(B) 627-633, (2014).
14. Raju C.N., Harikiran C., Priya T.S., Design of primary screening tool for early detection of breast cancer. Journal of Advances in Information Technology, 3(4):228–235, (2012).