# Design of Primary Screening Tool for Early Detection of Breast Cancer

Dr. C. Naga Raju
Professor and Head of IT, L.B.R.College of Engineering/Dept of IT, Mylavaram, India
Email: cnrcse@yahoo.com

C. Harikiran
Asst. Professor.
L.B.R.College of Engineering/Dept of IT, Mylavaram, India
Email: hari.ck10@gmail.com

T. Siva Priya
Asst. Professor.
L.B.R.College of Engineering/Dept of IT, Mylavaram, India
Email: sivapriya_tummala@yahoo.co.in

**Abstract—The innovative approach consists of using the same algorithmic core for processing images to detect both microcalcifications and masses. Despite the advancement in the medical sciences cancer is claiming more than 50% of the people afflicted by it every year. Of all cancer incidence women around the world, the most commonly diagnosed type of non-skin cancer which results in death is Breast Cancer and this can be best detected by digital mammography. This paper includes the design and development of software expert system for real time mammogram image analysis. The system so designed would give the radiologist an idea about the exact shape and size of any tumor present in the breast. Radiologists however are unable to detect the cancerous growth when benign though it is detected in the mammograms due to varying criteria like dense flesh around the cancer or distractions due to neighboring features. This problem will be resolved by using Digital Image Processing techniques like Image Segmentation where the image will be segmented into similar regions by meaningfully assigning class labels to similar pixels in a region. Hence the cancerous growth will be detected in its early stage and the Radiologists will be able to do better diagnosis because Image Segmentation techniques are simple yet very effective. In this paper an innovative method is applied which consist mainly three steps. In the first step normalizes the regions in the breast images through uniform distribution of histogram equalization. In the second step fuzzy logic is applied to remove ambiguity in the misclassification region and in the third step a new Weight is applied to the previously extended OTSU method.**

*Index Terms*—**Breast Cancer, Mammogram, ROI, OTSU Threshold, Histogram Weight, Mean, Variance**

## I. INTRODUCTION

Mammography and finding of suspicious masses during self-examinations form the primary screening tools for early detection of breast cancer [1, 2]. Early detection is difficult since the shape and size of the microcalcification clusters and speculated or irregular masses vary, and they are embedded in and camouflaged by various tissue structures [3, 4]. Mammography is the most effective method for the early detection of breast diseases. However, the typical diagnostic signs such as microcalcifications and masses are difficult to detect because mammograms are low-contrast images [5, 6].Breast cancer is considered as one of the primary causes of women mortality. The mortality rate in asymptotic women can be brought down with the aid of premature diagnosis. Despite the increasing number of cancers being diagnosed, the death rate has been reduced remarkably in the past decade due to the screening programs. Premature detection of breast cancer increases the prospect of survival whereas delayed diagnosis frequently confronts the patient to an unrecoverable stage and results in death [7, 8]. The Indian metropolises of Mumbai, Calcutta, and Bangalore display 23% of all the female cancers as breast cancers followed by cervical cancers [9, 10]. Despite the fact that the incidence of breast cancer in India is comparatively lower than that of the western countries, the issue is highly alarming. High quality images and mammographic interpretation are mandatory for the detection of premature and delicate symptoms of breast cancer. Mammogram (breast X-ray) is the medical image essential for the diagnosis of breast cancer and is considered to be the most dependable technique for premature detection. The widely recognized tool for the early detection of breast cancer in women


doi:10.4304/jait.3.4.228-235

with no symptoms; and to detect and diagnose breast disease in women experiencing symptoms like a lump, pain or nipple discharge, is mammography[11,12].

Contemporarily, screening mammography and radiographic imaging of the breast are the most effective tools for premature detection of breast cancer. Screening mammographic assessments are carried out on asymptomatic woman to detect premature, clinically unsuspected breast cancer. Still, studies have proved that all breast cancers that are retrospectively detected on the mammograms are not detected by radiologists. Due to the subtle and complex nature of the radiographic findings related with breast cancer, human factors such as varying decision criteria, distraction by other image features, and simple oversight can be responsible for the errors in radiological diagnosis. Computer assisted schemes that work on image processing and pattern recognition techniques can be utilized to enhance the diagnostic efficiency and for the location and classification of probable lesions and thereby alerting the radiologist to observe these areas with specific attention [13, 14]. Radiologists look out for particular abnormalities on mammograms visually. Some significant signs that radiologists pay attention to are clusters of microcalcifications, masses, and architectural deformations. A space-occupying lesion that is visible at more than one projection is referred to as a mass. Masses are illustrated with the aid of shape and margin features. Tiny deposits of calcium those are visible as minute bright spots on the mammogram are called as calcifications. They are exemplified by their type and distribution characteristics. The existence of microcalcifications is one of the significant and probably the only indication of cancer on a mammogram. A majority of the researches on computer analysis of mammograms have focused on the detection of small abnormalities, precisely the micro calcifications.

## II. OTSU METHOD

Otsu thresholding proposed a criterion for maximizing the between-class variance of pixel intensity to perform picture thresholding [15, 16]. Basic OTSU Thresholding technique involves segmenting or decomposing the entire image into regions of some similar properties like pixels of same intensities for further analysis. Hence using this method the image can be separated into dark and light regions. This is called as Thresholding the image. The separated regions are called assigned class labels where the intensity levels of each pixel in one region will be greater than the Threshold value and the intensity levels of pixels in the second region will be less than the Threshold value. The high frequency components in the resultant image are enriched whereas the low frequency background structure was removed. A global threshold value applied on the reconstructed image acquired for each mammogram and a binary image providing all the probable points of microcalcifications formed. The threshold value is automatically obtained from the grey level histogram with the application of a peak detection method

A significant technique for image segmentation that attempts to recognize and extract a target from its background with the aid of the distribution of gray levels or texture in image objects is referred to as Thresholding [17]. The statistics of the one-dimensional (1D) histogram of gray levels and on the two-dimensional (2D) co-occurrence matrix of an image form the basis of a majority of the thresholding techniques. Precisely, the discriminant criterion chooses the optimal threshold in order to maximize the separability of the resultant classes in gray levels. The procedure makes use of only the zeroth- and the first-order cumulative moments of the gray-level histogram and hence is trouble-free[18]. It is possible to extend the method to multithreshold problems in an uncomplicated manner.

### A. Methodology

An image is a 2D grayscale intensity function, and contains pixels with gray levels from 1 to L.
The probability of gray level in an image is:

$$P_i = f_i/N \Rightarrow \text{(number of pixels with gray level/total number of pixels)}$$

In the case of bi-level thresholding of an image; the pixels are divided into two classes, $C_1$ with gray levels [1, 2….t] and $C_2$ with gray levels [t+1….L].

Then, the gray level probability distributions for the two classes are

$$C_1 = p_1/w_1(t),\ldots p_t/w_1(t)$$
$$C_2 = p_{t+1}/w_2(t),\ p_{t+2}/w_2(t),\ldots,p_L/w_2(t)$$

$$\text{Where } w_1(t) = \sum p_i \text{ (where } i = 1, 2, 3..., t)$$
$$\text{and } w_2(t) = \sum p_i \text{ (where } i = t+1, t+2...L)$$

Also, the means for classes and are

$$\mu_1 = \sum ip_i/w_1(t) \text{ (where } i = 1, 2, 3..., t) \text{ and}$$
$$\mu_2 = \sum ip_i/w_2(t) \text{ (where } i = t+1, t+2... L)$$

Let $\mu_T$ be the mean intensity for the whole image. It is easy to show that

$$w_1\mu_1 + w_2\mu_2 = \mu_T \text{ and also } w_1+w_2 = 1$$

Otsu defined the between-class variance of the threshold image as

$$\sigma_B^2 = w_1(\mu_1 - \mu_T)^2 + w_2(\mu_2 - \mu_T)^2$$

Likewise the above formula can be extended for use in case of multiple thresholds extension and Proposed method for basic OTSU method.

## III. EXTENSION OF OTSU METHOD

The above OTSU method is simple and easier. However it fails if the Histogram is unimodal or close to unimodal. Hence an extension to the basic OTSU method will be implemented by selecting an optimal threshold. In this extended method the gray level distribution will be described using the average variance instead of average mean which is normally used in the basic OTSU method.

Here $\mu_1(t)$ and $\mu_2(t)$ can be regarded as the objects center gray and the background's center gray respectively, $\mu_T$ is the whole image center. This method makes sure that $(\mu_1 - \mu_2)^2$ is as bigger as it can get and gray distribution can be described not only by gray mean, but also by gray variance. The average variance will be used here to replace average mean in the basic OTSU method. The image variance reflects image uniformity; the variance is small inside of the objects and background. But the variance of edge and its neighborhood changes acutely. Hence it is reasonable to use average variance instead of the foreground and the background means in OTSU method.

$$t^* = \text{Arg Max}[w_1(\sigma_1^2(t) - \sigma_T^2(t))^2 + w_2(\sigma_2^2(t) - \sigma_T^2(t))^2$$

$$\sigma_0^2(t) = 1/w_0(t) \sum (i - \mu_1(t))^2 \, p(i) \text{ (where } i = 1, 2, 3..., t)$$
$$\sigma_1^2(t) = 1/w_1(t) \sum (i - \mu_1(t))^2 \, p(i) \text{ (where } i = t+1,.., m-1)$$
$$\sigma^2(t) = \sum (i - \mu_T(t))^2 \, p(i) \text{ (where } i = t+1, t+2, ..., L)$$

This method represents well adaptability and certain anti-noise abilities; it will not be although this method has some difficulties processing images with unimodal distribution.

## IV. NOVEL FUZZY OTSU METHOD

**Step1:** Features are based on the grey-level histograms from selected regions of the breast. The distances to the skin normalized from 0 to 100 (providing invariance to the size of the breast) are utilized in the construction of the regions. Histogram modeling techniques alter an image in order to ensure that the histogram is of the desired shape. This is beneficial for the elongation of low levels of mammograms with the narrow histograms. Histogram equalization is a conventional histogram modeling methodology. According to the information theory, the uniform distribution attains maximum entropy, which encloses the most information. Thus, the mammogram information needs to be maximized in order to redistribute the gray-levels and achieve at the most uniform histogram .The next is fuzzy logic which produce optimal threshold to avoid the fuzziness in the image and makes good regions regarding background and object.

**Step2:** Fuzzy set theory assigns a membership degree to all elements among the universe of discourse according to their potential to fit in some class. The membership degree can be expressed by a mathematical function $\mu_A(x_i)$ that assigns, to each element in the set, a membership degree between 0 and 1.Let be the universe (finite and not empty) of discourse and $x_i$ an element of .A fuzzy set **A** in **X** is defined as

$$A = \{(x_i, \mu_A(x_i)) | x_i \in X \}$$

The S –function is used for modeling the membership degrees. This type of function is suitable to represent the set of bright pixels and is defined as

$$\mu_{AS}(x) = s(x, a, b, c)$$
$$= \begin{cases} 0, & x \le a \\ 2\left\{\dfrac{(x-a)}{(c-a)}\right\}^2, & a \le x \le b \\ 1 - 2\left\{\dfrac{(x-c)}{(c-a)}\right\}^2, & b \le x \le c \\ 1, & x \ge c \end{cases}$$

Where $b = (1/2)(a + c)$

The S–function show in the Fig1 can be controlled through parameters **a** and **c**. Parameter **b** is called the cross over point where $\mu_{AS}(b) = 0.5$. The higher the gray level of a pixel (closer to white), the higher membership value and vice versa.

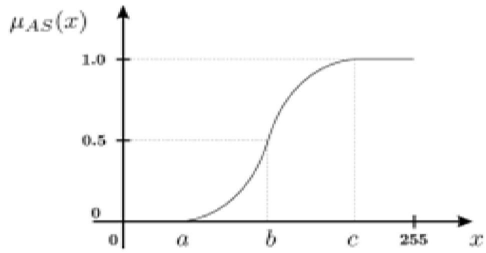



Fig1.Typical shape of the S-function function

Measures of Fuzziness are **a** reasonable approach to estimate the average ambiguity in fuzzy sets is measuring its fuzziness. The fuzziness of a crisp set should be zero, as there is no ambiguity about whether an element belongs to this t or not. If $\mu_A(x) = 0.5, \forall x,$ the set is maximally ambiguous and its fuzziness should be maximum. Degrees of membership near 0 or 1 indicate lower fuzziness, as the ambiguity decreases. Kaufmann in introduced an index of fuzziness (IF) comparing a fuzzy set with its nearest crispset. A fuzzy set **A*** is called crispset of **A** if the following conditions are satisfied:

$$\mu_A(x) = \begin{cases} 0, & \text{if } \mu_A(x) < 0.5 \\ 1, & \text{if } \mu_A(x) \ge 0.5 \end{cases}$$

**Step3:** The optimal threshold value exists at the valley of the two peaks or at the bottom rim of a single peak. The valley in the histogram that separates the object from the background, its probability of occurrence is small in gray level histogram. Because of the optical threshold should near the cross where the object and the background intersect. The probability of occurrence at the threshold value should divide into two parts. Its first half belongs to background and second half belongs to object. Then we apply a new weight M to the OTSU method.

$$t = (P1*D1+ P2*D2)_{*} M$$

$$\text{where } D1=(\sigma_1^2(o)- \sigma_T^2(t))^2$$

$$D2=(\sigma_2^2(b)- \sigma_T^2(t))^2$$

$$M=(1-P_T(t)/2)$$

Using this method we can make sure that the result threshold value resides at the valley or at the bottom of the right rim of single peak. It's also maximizes group variance and ensures that both the variance of the object and that of the background keep away from the variance of the whole image. Smaller the p(t)/2, larger will be the weight.

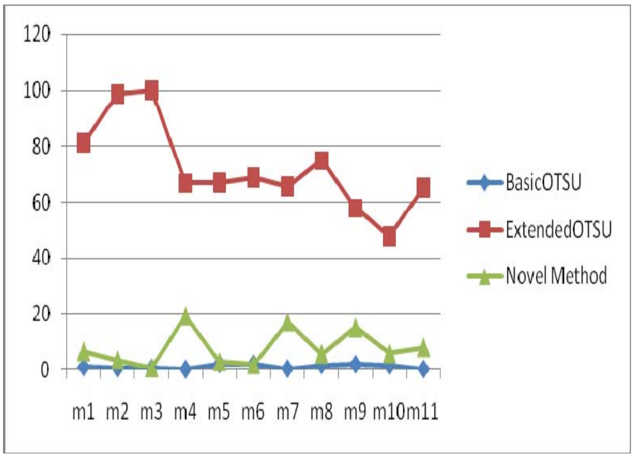
## V. EXPERIMENTAL RESULTS

In order to verify the effectiveness of the segmentation process using the proposed method, a set of images of different kinds were tested. Experimental results illustrate that the system is capable of aiding the interpretation of radiologists in their daily practice besides enhancing their diagnostic performance. The performance evaluation of three methods have been described based on table of values and graph shown in the paper .Results of grap1 reflect the severity of cancer using the above three methods. From the Basic OTSU method we can infer that the values plotted in graph1 are too low, hence we cannot clearly differentiate among the normal, moderate or severely cancer affected breast. The Extended OTSU method produced better results than basic OSTU method. But from the Extended OTSU method we can infer that the values plotted in graph1 are too high and ambiguous values among some of breast images, hence we cannot clearly differentiate among the normal, moderate or severely cancer affected breast. From the Novel Method we can infer that the values plotted in graph1 are neither too low nor too high (i.e., values are moderate). Further comparing the plotted values in graph1 against each of the images in table1 it is evident that the values clearly reflect the levels of severity of the cancer. For example the plotted values for images M4, M7, and M9 show that the severity of cancer in these respective images is too high, while the values for images M1, M8, M10, M11 show that the severity of cancer in these respective images is moderate and the values for images M2, M3, M5, M6 show that the severity of cancer in these respective images is low. Using the first two methods the levels of severity of cancer in respective images is not clearly reflected in the graph1. This is due to the fact that the values for Basic OTSU and Extended OTSU methods from table1 are either too low or too high. In the proposed method the defect can be extracted more precisely. It's able to select optimal threshold values for both unimodal and bimodal distribution, so it can be used on various defect detection mammograms applications.

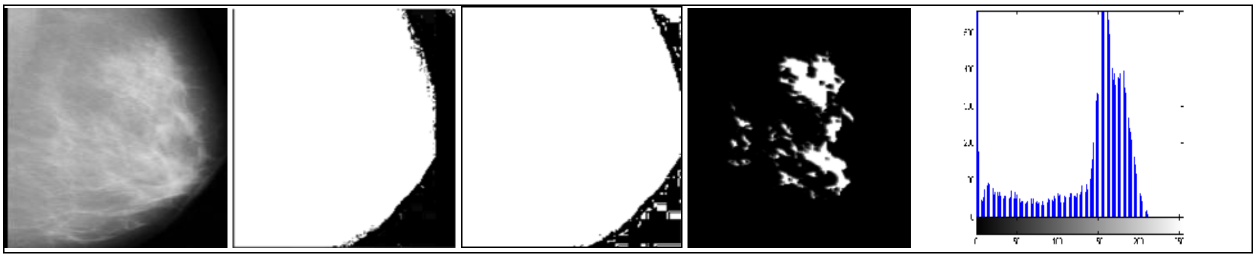
TABLE I.

| Images | BasicOTSU | ExtendedOTSU | Novel Method |
|---|---|---|---|
| M1 | 0.88 | 81.04 | 6.81 |
| M2 | 0.47 | 98.65 | 3.71 |
| M3 | 0.56 | 100 | 0.85 |
| M4 | 0.13 | 66.57 | 19.17 |
| M5 | 1.99 | 66.7 | 3.06 |
| M6 | 2.03 | 68.49 | 2.04 |
| M7 | 0.31 | 65.49 | 17.09 |
| M8 | 1.69 | 74.67 | 5.84 |
| M9 | 1.84 | 57.57 | 15.09 |
| M10 | 1.47 | 47.45 | 6.04 |
| M11 | 0.21 | 65.01 | 8.21 |

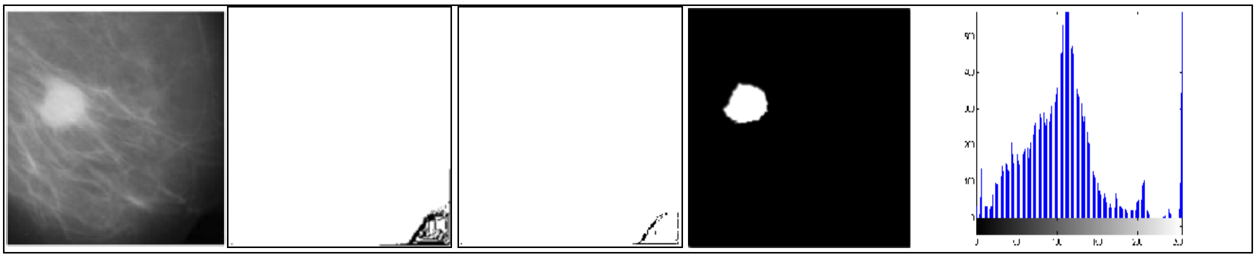
GRAPH1



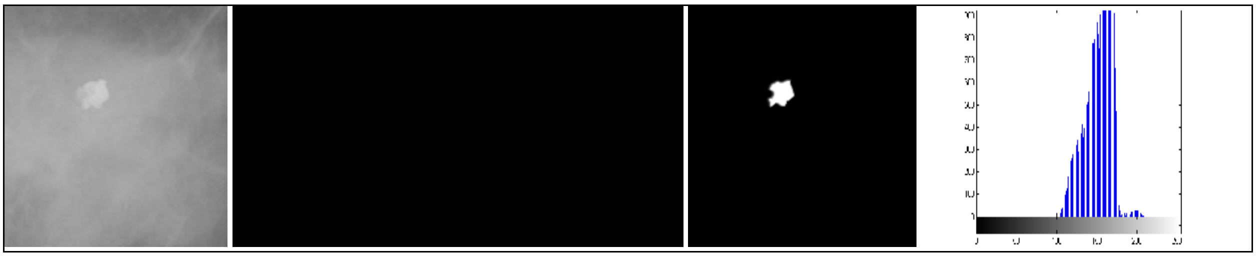
RESULTS



M1 OTSU Extended OTSU Novel Fuzzy OTSU m1

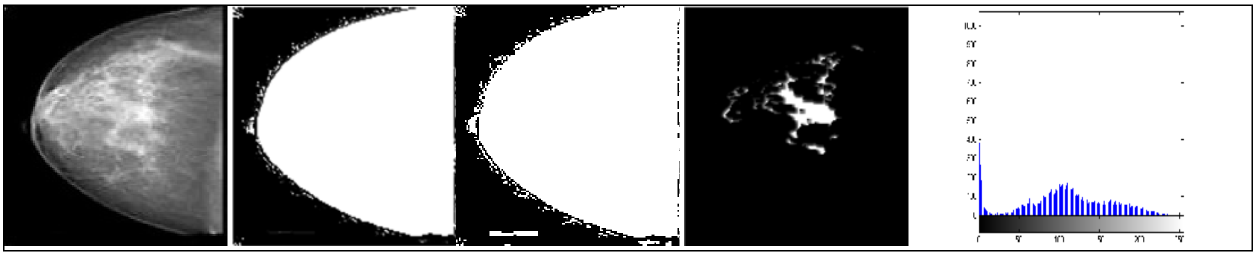


M2 OTSU Extended OTSU Novel Fuzzy OTSU m2

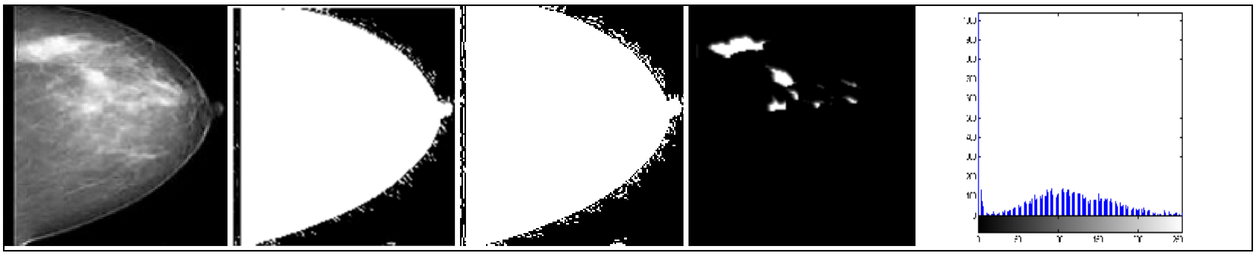


M3 OTSU Extended OTSU Novel Fuzzy OTSU m3

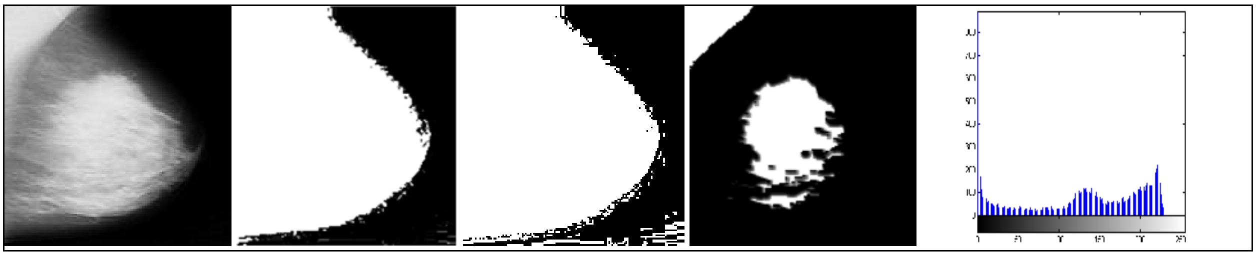
M4 OTSU Extended OTSU Novel Fuzzy OTSU m4



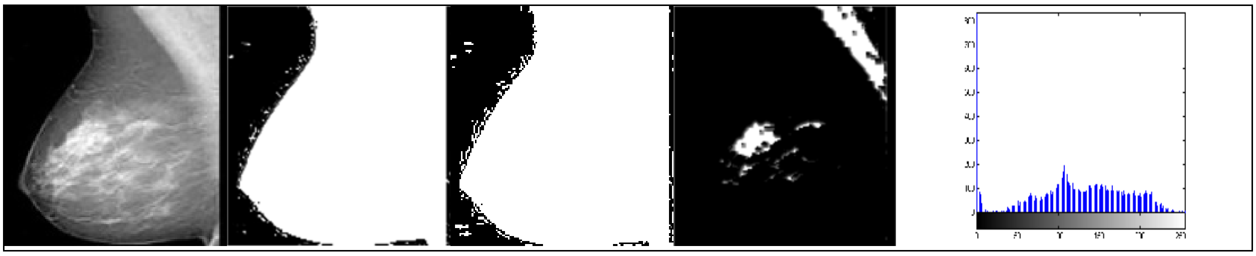
M5 OTSU Extended OTSU Novel Fuzzy OTSU m5



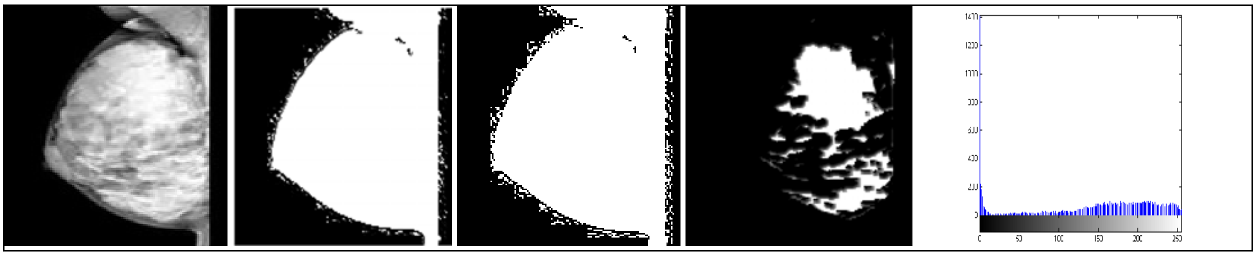
M6 OTSU Extended OTSU Novel Fuzzy OTSU m6



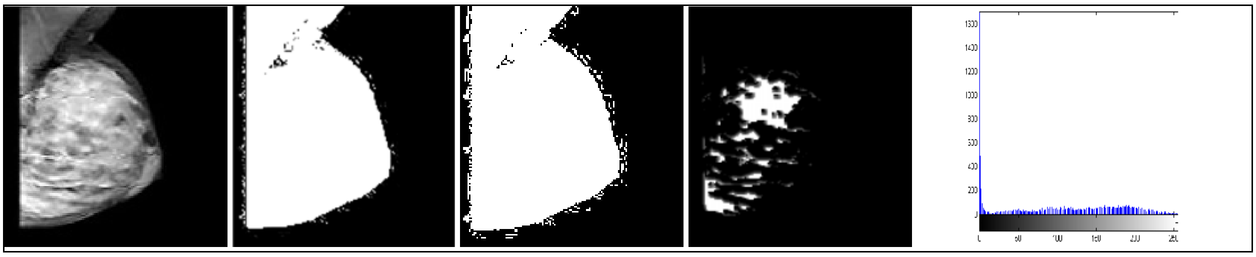
M7 OTSU Extended OTSU Novel Fuzzy OTSU m7



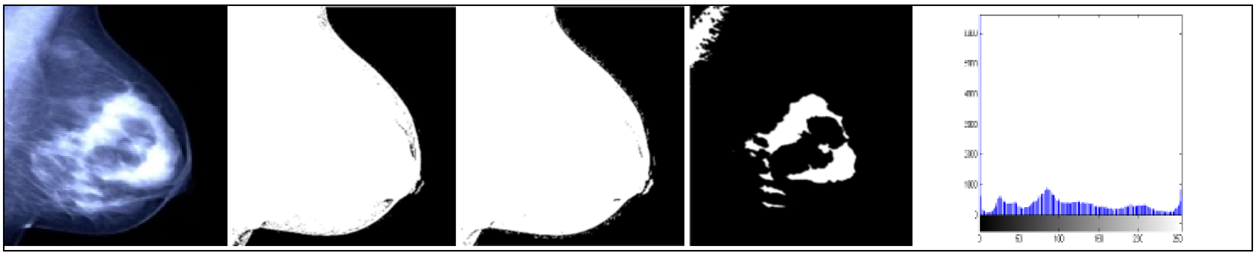
M8 OTSU Extended OTSU Novel Fuzzy OTSU m8



M9 OTSU Extended OTSU Novel Fuzzy OTSU m9



M10 OTSU Extended OTSU Novel Fuzzy OTSU m10



M11 OTSU Extended OTSU Novel Fuzzy OTSU m11

## VI. CONCLUSION

The contemporary preference for the premature detection of breast cancer in women is Mammography. The elucidation of mammograms greatly depends on radiologist's opinion. The approach depends on an OTSU threshold operator strategy, for the segmentation of mass/microcalcification. In the proposed work we have assessed an automated detection method for one of the principal signs of breast cancer: clusters of microcalcifications and mass lesions. This technique involves normalizing regions of breasts and thresholding the Region of Interest (ROI) by using basic OTSU method. This method however fails if the Histogram is unimodal or close to unimodal. Hence an extension to the basic OTSU method will be implemented by selecting an optimal threshold. In this extended method the gray level distribution will be described using the average variance instead of average mean which is normally used in the basic OTSU method. So the average variance will replace average mean in the extended method. Furthermore extending this technique, a new method is proposed. In this method fuzzy logic is applied to remove ambiguity in the misclassification region and a new Weight is applied to the previously extended OTSU method. This proposed method ensures that both the variance of the object and variance of background are far from the variance of the image. This Weight ensures that the threshold is optimal and we will get satisfactory results for images with histogram of unimodal or multimodal distribution.

## REFERENCES


[1] R.A. Smith, "Epidemiology of breast cancer in a categorical course in physics," Technical Aspects of Breast Imaging, 2nd ed., RSNA publication, Oak Book, II, pp.21, 1993.

[2] R. Peto, J. Boreham, M. Clarke, C. Davies., V. Beral, "UK and USA Breast cancer deaths down 25% in year 2000 at ages 20-69 years", THE LANCET, Volume 355, Issue 9217, Page 1822, 20 May 2000.

[3] Ranadhir Ghosh, Moumita Ghosh, John Yearwood, "A Modular Framework for Multi category feature selection in Digital mammography", In Proceedings of the 12th European Symposium On Artificial Neural Networks ESANN'2004, Bruges (Belgium), pp. 175-180, 28-30 April 2004.

[4] Bernard W. Stewart, Paul Kleihues, "WORLD CANCER REPORT", WHO, International Agency for Research on Cancer, IARC Press, Lyon 2003.

[5] E. L. Thurfjell, K. A. Lernevall, and A. A. Taube, "Benefit of independent double reading in a population-based mammography screening program," Radiology, vol. 191, pp. 241–244, 1994.

[6] J. G. Elmore, C. K. Wells, C. H. Lee, D. H. Howard, and A. R. Feinstein, "Variability in radiologists' interpretations of mammograms," The New England journal of Medicine, vol. 331, no. 22, pp. 1493–1499, 1994.

[7] Huai Li, K. J. Ray Liu, and Shih-Chung B. Lo, "Fractal Modeling and Segmentation for the Enhancement of Microcalcifications in Digital Mammograms", IEEE Transactions on Medical Imaging, vol. 16, no. 6, December 1997.

[8] A. S. Pednekar and I. A. Kakadiaris, "Image segmentation based on fuzzy connectedness using dynamic weights," IEEE Trans. Image Process., vol. 15, no. 6, pp. 1555–1562, Jun. 2006.

[9] S. Sahaphong and N. Hiransakolwong, "Unsupervised image segmentation using automated fuzzy c-means," in Proc. IEEE Int. Conf. Computer and Information Technology, Oct. 2007, pp. 690–694.

[10] Veldkamp, W., Karssemeijer, N., "Accurate segmentation and contrast measurement of microcalcifications in mammograms: A phantom study", Medical Physics, vol. 25, pp. 1102-1110, 1998.

[11] L. Shen, R. Rangayyan, and J. Desaultels, "Detection and Classification Mammographic Calcifications", International Journal of Pattern Recognition and Artificial Intelligence. Singapore: World Scientific, pp. 1403–1416, 1994.

[12] K. Bowyer and S. Astley, "The Art of Digital Mammographic Image", Singapore: World Scientific, vol. 7, 1994.

[13] H. Barman, G. Granlund, and L. Haglund, "Feature extraction for computer- aided analysis of mammograms," in State of the Art of Digital Mammographic Image Analysis. Singapore: World Scientific, 1994, vol. 7, pp. 128–147.

[14] N. Otsu, A threshold selection method from gray-level histogram, IEEE Transactions on Systems Man Cybernet, SMC-8 pp. 62-66, 1978.

[15] H.F. Ng, "Automatic thresholding for defect detection", Pattern Recognition Letters, (27): 1644-1649, 2006.

[16] M. Sezgin and B. Sankur, Survey over image thresholding techniques and quantitative performance evaluation, Journal of Electronic Imaging, pp.146-156, 2003

[17] H. Lee and R. H. Park, Comments on an optimal threshold scheme for image segmentation, IEEE Trans. Syst.Man Cybern, SMC-20,741-742, 1990
[18] J. Z. Liu and W. Q. Li, The Automatic thresholding of gray-level picture via two-dimensional Otsu method, Acta Automatica Si.19, 101-105,1993

**Dr. C. Naga Raju** received his B.Tech degree in Computer Science from J.N.T.University Anantapur, M.Tech degree in Computer Science from J.N.T.University Hyderabad and PhD in digital Image processing from J.N.T.University Hyderabad. Currently, he is working as a professor & Head of IT in LakiReddy Bali reddy College of Engineering, Vijayawada. He is professor incharge for systems department. He has got 15 years of teaching experience. He has published thirty research papers in various national and international journals and about twenty eight research papers in various national and international conferences. He has attended twenty seminars and workshops. He is member of various professional societies like IEEE, ISTE and CSI.

**Mr. C. Harikiran.** He was born in Vijayawada, Andhra Pradesh, India. He received B.E (Bachelor of Engineering) degree in CSE, Computer Science & Engineering from University of Madras, Masters Degree M.Tech in Software Engineering Jawaharlal Nehru Technological University, Kakinada. He worked for over nine years as an IT Consultant in New York City. Presently he is working as Asst. Professor in Lakireddy Bali Reddy College of Engineering, Vijayawada. He plans to pursue PhD. His research interests are in the areas of Digital Image Processing, Artificial Intelligence, and Neural Networks, Information Security.

**Siva Priya Tummala** received her B.Tech degree in Information Technology from P.V.P S.I.T Vijayawada. She has got 2 years of teaching experience from Paladugu Parvathi Devi Institute of Engineering And Technology. Currently, she is working as Assistant Professor at Lakiredy Baliredy College of Engineering, Mylavaram. She has published two research papers in international journal.