Statistical Based Texture Feature extraction to deduct and extract the tumor in Mammogram image.

M Punitha[1] and K.Perumal[2]

[1]*Research scholar, Department of Computer Applications, Madurai Kamaraj University, Madurai.*
Punitha_2883@rediff.com

[2]*Professor, Department of Computer Applications, Madurai Kamaraj University, Madurai.*
perumalmkucs@gmail.com

ABSTRACT

Currently digital mammography is the most efficient and widely used technology for early breast cancer detection. The key factors in diagnosing digital mammogram does not provide accurate results and hence we proposed an improved statistical based feature extraction technique and shape features is proposed here to deduct and extract the tumor in mammogram image. This SBFE have three types of statistics to find the tumor levels. There are FOS (first order statistics) SOS (Second order statistics) and HOS (Higher order statistics) thus, the dimension of data features is reduced and hence the classification accuracy rate is improved. The quick reduct algorithm is applied for the HOS and it helps to find the tumor with in short time and also Shape features to identify the tumor shape as well. The result proves the efficacy of the proposed method in classifying tumor easily and predictive accuracy of the image classified the accuracy of the mammogram image.

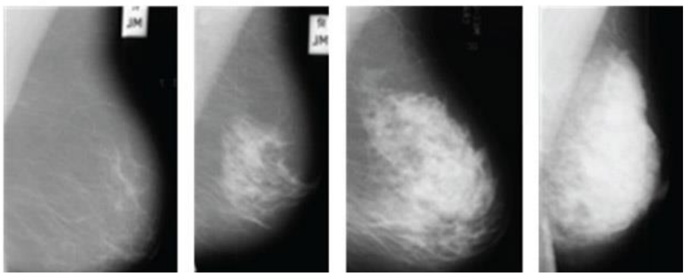
KEYWORDS
Image segmentation, Shapes, feature extraction techniques, and quick reduct algorithm, SBFE and Predictive accuracy.

I. INTRODUCTION

In recent decades, Digital mammography is one of the most suitable methods for early detection of breast cancer. Early detection can play an effective Role in the prevention. [1] However, it is difficult for radiologists to give both exact and uniform assessment for the colossal number of Mammogram created in far reaching screening. There are some limitations for human observers as some anomalies may be missed due to human Error. Digital Mammograms allow manipulation of fine differences in image contrast by means of image processing algorithms. The effectiveness of digital mammography in detection of breast cancer is currently under investigation. A variety of algorithms have been developed by independent investigators for use with digital mammograms. So a new approach called SBFE is to extract the tumor easily in various processing levels.

II. MAMMOGRAM IMAGE DATASET

Images from MINI MIAS (Mammographic Image Analysis Society) data base are used for the evaluation of the system. In the mini MIAS data base the original MIAS has been taken. The sample images of two different normal and abnormal samples are shown at Figure 1.



(a) (b) (c) (d)

Figure1: Normal and Abnormal imaging samples: (a) Normal (b) First stage or starting stage (c) Second stage (d) Higher stage of tumor portion. Images are analyzed from this dataset.

III. PROPOSED METHODOLOGY

The Statistical based feature extraction techniques should be processed in various levels and analyzed with segmentation processed

Phase 1: The pre-processing steps (i.e. Gray scale conversion, Image Histogram)

Phase 2: The results of SBFE has been processed with the segmentation and classification and getting results as normal or abnormal by using quick reduct algorithm.

Phase 3: Predictive accuracy of the tumor are found out in all input images.

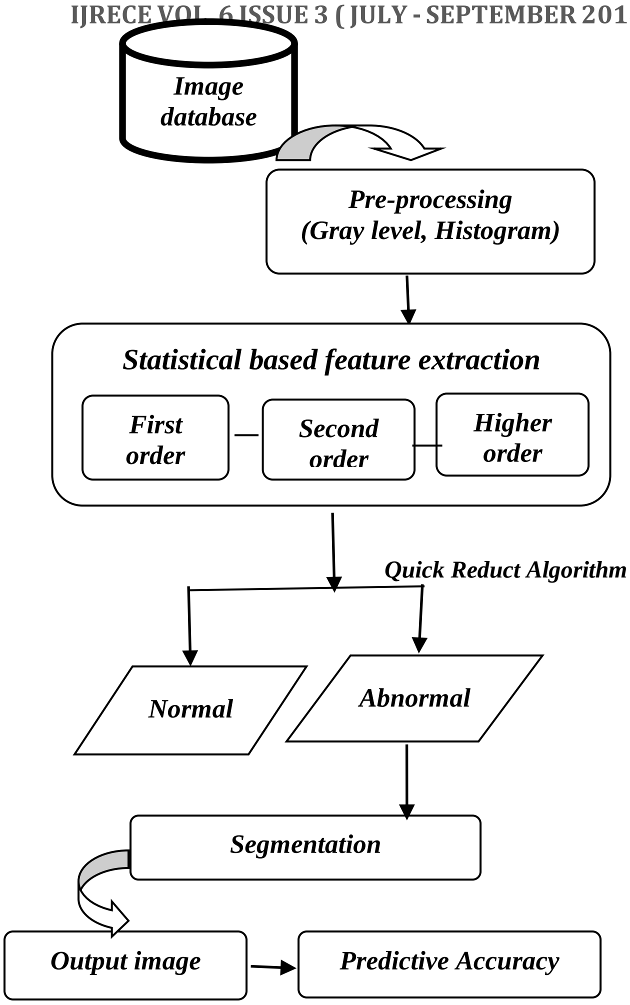


Figure 2: Flow diagram for proposed SBFE

3.1 Image Preprocessing

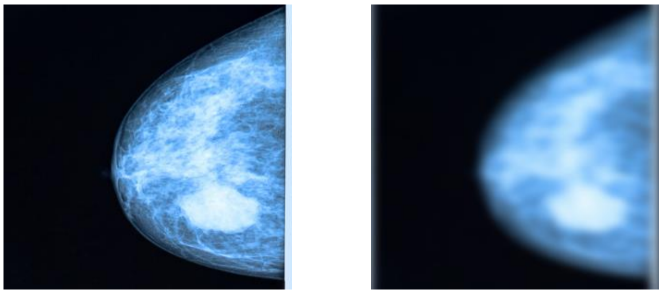
The pre-processing of mammogram image is essential before segmentation of the tumor. In preprocessing of the mammogram images Noise reduction and Image segmentation are done in different techniques. the objective of this process is to improve the quality of the image to make it ready for further processing by removing the irrelevant and unwanted parts in the background of the mammogram.

3.2 Database

To evaluate the proposed method, the Database from Screening Mammography database is used for the experiment. The DDSM cancer dataset was obtained from a university of south Florida.
Images are available online at the
http://marathon.csee.usf.edu/Mammography/DDSM

3.3 Gray level

The Gray level has been applied to improve the contrast of the image. We have used for prepressed to improve the image quality. This process can be achieved by adjusting the grey level and dynamic range of the image, which is the deviation between minimum and maximum pixel value. Similarly as with 8-bit grayscale, the lightness of the gray is specifically relative to the number to representing the brightness levels of the primary colors. In a grayscale image, the hue and saturation of each pixel is equal to zero.

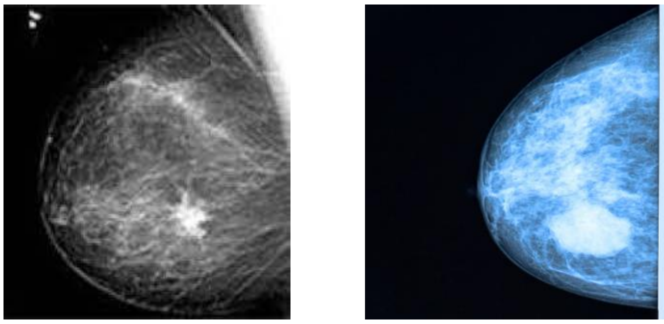


(a)   (b)

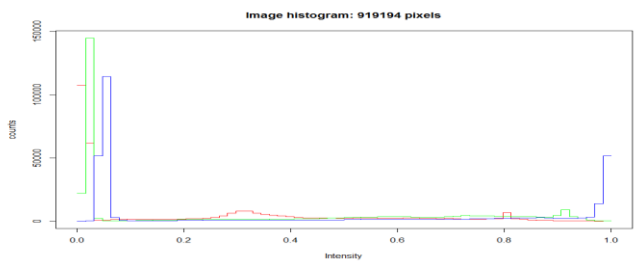
Figure 3: (a) Original image (b) Gray level of the image.

3.4 Image Histogram

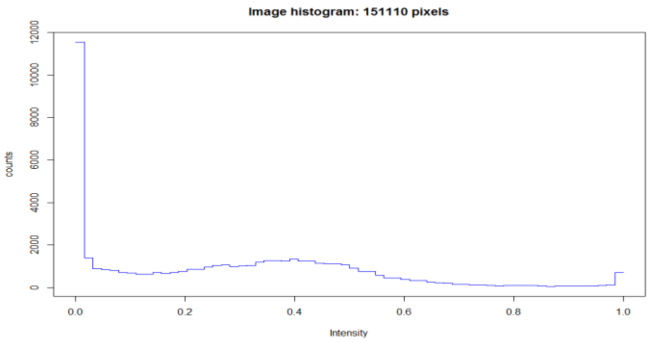
Histogram equalization is used for contrast adjustment using the image histogram. The histogram of an image refers to intensity values of pixels. The Histogram shows the number of pixels in an image at each intensity value. When ROI is represented by close contrast values, this histogram equalization enhances the image by increasing the global contrast.



(a) original image   (b) original image



(c) Histogram of the original image a



(d) Histogram of the original image b

Figure 4: (a) and (b) original image and (c) and (d) histogram of the images

Figure 4: shows the image histogram of an input image and it shows the distribution of pixels among
Those grayscale values. The 8-bit gray scale image is having 256possible intensity values. A narrow histogram indicates the low contrast region.

IV. TEXTURE FEATURE EXTRACTION

Neville et al, discussed texture features can be extracted using several methods such as statistical, structural, and model-based and transform Information. Here statistical based feature extraction methods is proposed.

4.1 Statistical based Feature Extraction

Statistical methods characterize the texture indirectly according to the non-deterministic properties that manage the relationships between the gray levels of an image. Statistical methods are used to analyze the spatial distribution of gray values by computing local features at each point in the image and deriving a set of statistics from the distributions of the local features. The statistical methods can be classified into first order (one pixel), second order (pair of pixels) and higher order (three or more pixels) statistics. The first order statistics estimate the properties (e.g. average and variance) of individual pixel values by waiving the spatial interaction between image pixels. The second order and higher order statistics estimate properties of two or more pixel values occurring at specific locations relative to each other. The Most popular second order statistical features for texture analysis are used here.

4.2 First Order Statistics

First Order histogram provides different statistical properties and focused on four statistical moments of the intensity histogram of an image. These depend only on individual pixel values and not on the interaction or co-occurrence of neighboring pixel values. The four first order histogram statistics are mean, variance, skewness and kurtosis. It's depends only on individual pixel values and not on the interaction or co-occurrence of a histogram *h* for a gray scale image *I* with intensity values in the

Range *I* (*x*, *y*) 0, *K* 1 would contain exactly *K* entries, where for a typical 8-bit grayscale image, *K* 28 256. Each individual histogram entry is defined as, h (*i*) = the number of pixels in *I* with the intensity value *I* for all 0 *i K* .

The Equation (1) defines the histogram as,

h (i) =*cardinality {(x, y) | I(x, y) i}* ………..*1*

Where, *cardinality* denotes the number of elements in a set. The standard deviation, and skewness of the intensity histogram are defined in Equation (2) and (3).

$$\sum (I(x, y) - m)^2 / n \qquad \ldots\ldots\ldots..2$$

$$\text{Skewness}=\sum (I(x, y) - m)^3 / n^3 \quad \ldots\ldots3$$

The most frequently used central moments are Variance, Skewness and Kurtosis given by μ_2, μ_3, and μ_4 respectively. The Variance is a measure of the histogram width that measures the deviation of gray levels from the Mean. Skewness is a measure of the degree of histogram asymmetry around the Mean and Kurtosis is a measure of the histogram sharpness. So the tumor portion is partially identified for the given mammogram image.

4.3 Second Order Statistics

Here, the GLCM Method is used to extract the second order texture information from the image. The number of rows and columns are equal to the pixel values of the image. In this GLCM matrix compared the one image gray level to another gray level of the given image. The calculation use the content of the GLCM to given measures of the pixel values but computing the co-occurrence matrix on two parameters as well as the neibourhood pixels, that are the relative distance between the pixel values and measures values and consider to θ normally θ is quantized in four directions (eg.$0^{0,}$ 45^0 ,90^0 and 135^0) and also Various combination is possible, so the tumor portion can be calculated easily and identified the possible tumor portion exactly identified.

4.4 Higher order Statistics

This method helps to deduct the tumor portion in short period. Gaussian function used in this systems and extract the tumor due to deviation from Gaussianity and deduct the tumor by using nonlinear function. In this HOS we have to apply the quick reduct algorithm which helps to find the tumor easily.

- Quick Reduct Algorithm

Input: image, the image pixels A

Output: R, the attribute reduct, the tumor portion pixel reduct $R \subseteq$ image

(1) $R = \Phi$
(2) Do
(3) $T = R$
(4) For each $x \in$ (image − R)
(5) If ( ) ( ) γ RU{x} A > γ T A
(6) T=RU{x}
(7) End for
(8) R=T
(9) Until (A) (A) R C $\gamma = \gamma$
(10) Return R

QuickReduct algorithm initially starts with an empty points of pixel and includes an attribute in an iteration that increases the kappa in a maximum way. So the tumor may found quickly.

V. Predictive Accuracy.

Here we find the tumor for positive or negative that is PPV(positive pixel value) and NPV (negative pixel value) results in statistics and diagnostic tests that are true positive and true negative results, respectively. A high result can be interpreted as indicating the accuracy of the mammogram image.
By using this formula to predicting the value easily.

$$\text{Predictive value} = \frac{\text{Original Value} - \text{Predictive Value}}{\text{Predictive value}}$$

Each image is predicted as well, and also the accuracy of the tumor positions found out, by using this formula we can easily identify the tumor exactly and calculate the pixel values.

Table 1: Number of pixels and tumor predictive area

Number of the pixels and tumor area		
original image	Total Tumor pixel accuracy	predictive value accuracy
Image 1	0.36720	0.85000
Image 2	0.21000	0.72000
Image 3	0.33000	0.81100
Image 4	0.27840	0.72556
Image 5	0.24708	0.68452
Image 6	0.21176	0.61025

Table 1 represents the predictive accuracy of the given images very clearly and also found tumor portions for the given images.

Then the below mentioned graphical representation of predictive value pixel and calculated by original value. And the bar shows the pixel range of the given images and it may defer from image to image.

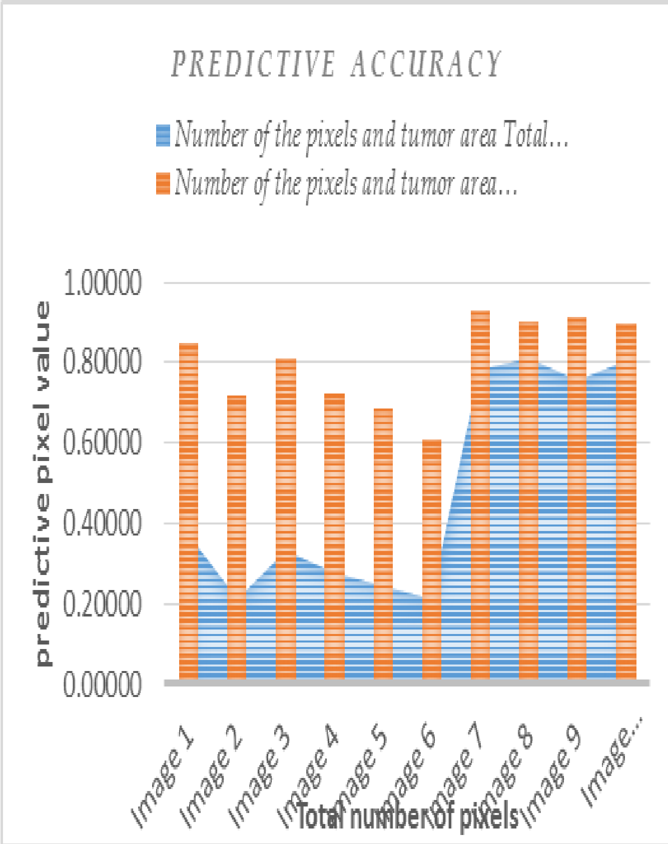


Figure 5: Predictive Accuracy Results for Various Textures Feature Extraction Methods

VI. RESULTS AND DISCUSSION

The collected Mammogram images are pre- processed and applied the Statistical based feature extraction by using the Quick reduct algorithm technique. The selection of size and shape of structuring element of an image I(x, y) depends upon the qualitative Selection of structural element size for one image is varied from another image and calculate the tumor portion calculated easily and accurately with in short time.

6.1 DISCUSSION

The main objective of work is to classify the Mammogram as normal and abnormal images with required features. The various texture classification analysis are investigated to differentiate normal and abnormal subjects of mammogram images. And also the proposed methods are applied with quick reduct algorithm. The predictive value are calculated for all images, the tumor image are extracted easily and getting clear image.

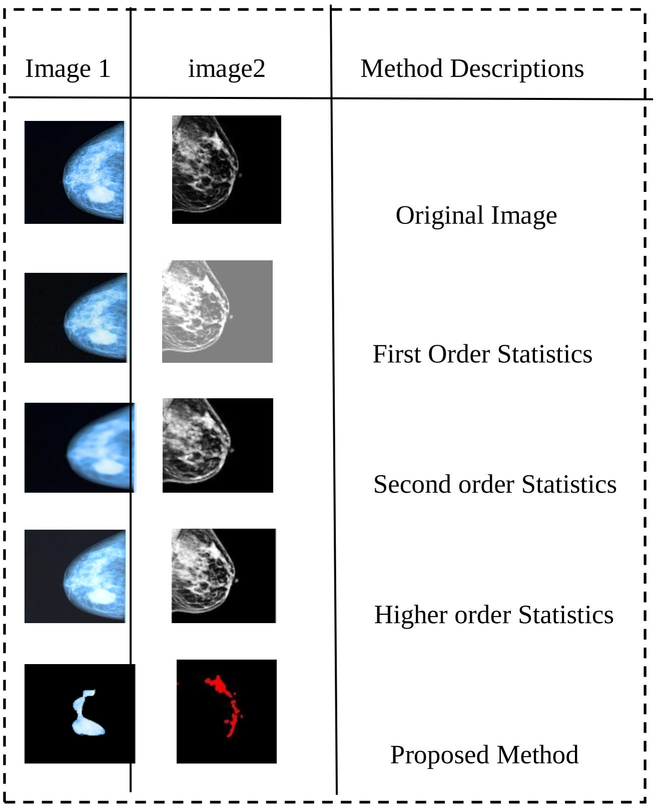


Figure 6: Results for Various image Textures Feature Extraction Methods

VII. CONCLUSIONS

The results obtained using Statistical based Feature Extraction approach is applied in different mammogram tumor images. And with various ordering levels. The Quality assessment of this method also compared with the other kind of techniques such as predictive accuracy is calculated for every images. And the quick reduct algorithm estimated and proved from the statistical assessments of this work gives the better results. This approach has good performance by removing noise, perceptual quality, and preserving the edge details without blurring in all images. Such Combination of this technique gives the better images.

References

[1] F. Moayedi, Z. Azimifar, R. Boostani, S. Katebi, Contourlet-based mammography mass classification, ICIAR 2007, Lecture Notes in Computer Science 4633, pp. 923–934, (2007)

[2] R. Highnam and M. Brady, Mammographic Image Analysis , Kluwer Academic Publishers, 1999. ISBN: 0-7923- 5620-9.

[3] Zaheeruddin, Z. A. Jaffery and Laxman Singh, "Detection and Shape Feature Extraction of Breast Tumor in Mammograms", Proceedings of the World Congress on Engineering 2012 Volume 2 WCE 2012, July 4 - 6, 2012.

[4] Moumena Al-Bayati and Ali El-Zaart, "Mammogram Images Thresholding for Breast Cancer Detection Using Different Thresholding Methods", Advances in Breast Cancer Research, 2, 72-77, 2013.

[5] K.Vennila, k.Sivakami and R.Padmapriya, "Detection of Mass in Digital Mammograms", International Journal of Computer Applications (0975 – 8887) Volume 104 – No.5, October 2014.

[6] Alfonso RojasDomínguez and Asoke K. Nandi, "Toward breast cancer diagnosis based on automated segmentation of masses in mammograms", Pattern Recognition, Vol. 42, pp. 1138 – 1148, 2009.

[7] Aminah Abdul Malek, Wan Eny Zarina Wan Abdul Rahmana, Arsmah Ibrahima, Rozi Mahmudb, Siti Salmah Yasirana and Abdul Kadir Jumaata, "Region and Boundary Segmentation of Microcalcifications using Seed-Based Region Growing and Mathematical Morphology", Procedia Social and Behavioral Sciences, Vol. 8, pp. 634–639, 2010.

[8] Dubeya R. B., Hanmandlub M. and Guptac S.K., "A comparison of two methods for the segmentation of masses in the digital mammograms", Computerized Medical Imaging and Graphics, Vol. 34, pp. 185–191, 2010.

[9] Wenda He, Erika R.E. Denton, Kirsten Stafford and Reyer Zwiggelaar, "Mammographic image segmentation and risk classification based on mammographic parenchymal patterns and geometric moments", Biomedical Signal Processing and Control, Vol. 6, pp. 321– 329, 2011.

[10] Usha R, Perumal K, An Automated Diagnosis of Magnetic Resonance Images for Brain Tumor using Stationary Wavelet based SFTA Texture Features, International Journal of Control Theory and Applications (IJCTA), ISSN: 0974-5572, 10(21) (2017), 17-24.

[11] Suman Thapur, Shevani Garg, Study and Implementation of Various Morphology Based Image Contrast Enhancement Techniques, International Journal of Computing and Business Research, Proceedings of I-Soceity 2012, (2012), pp.2229-6166.

[12] K.S. Srinivasan and D. Ebenezer, "A new fast and efficient decision based algorithm for removal of high density impulse noise", IEEE Signal Process. Lett., Volume 14, no. 3, pp.189-192, Mar. 2007.

[13]Jim Z.C. Lai, Yi-Ching Liaw, and Julie Liu, "A fast VQ codebook generation algorithm using codeword displacement", *Pattern Recogn.* vol. 41, no. 1, pp.: 315–319, 2008.

[14]C.H. Hsieh, J.C. Tsai, Lossless compression of VQ index with search order coding, IEEE Trans. Image Process. vol. 5, No. 11, pp.: 1579–1582, 1996.

[15]H. B. Kekre, Tanuja K. Sarode, Bhakti Raul, "Color Image Segmentation using Kekre's Algorithm for Vector Quantization", International Journal of Computer Science (IJCS), Vol. 3, No. 4, pp.: 287-292, Fall 2008. Available: http://www.waset.org/ijcs.

Authors

Short Biography

Mrs. M.Punitha is pursuing Ph.D in Madurai Kamaraj University, Madurai. She received the M.Phil Degree in Computer Applications from Madurai Kamaraj University in the year 2017. MCA degree from Fatima College Madurai in the year 2006. She has contributed papers in International Journals and Conferences. Her research interest is in image processing and image segmentation in mammogram images.

Mr. K. Perumal working as a professor in Madurai Kamaraj University, Madurai. He has contributed more than 25 papers in International Journals and Conferences. He has guiding 8 scholars. His interest includes Data Mining, Image processing and medical images.