ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2014

Detection of Breast Cancer Using Artificial Neural Networks

Anu Alias [1], B.Paulchamy [2]

PG Scholar, Department of ECE, Hindusthan Institute of Technology. Coimbatore,Tamil Nadu,India [1]

Assistant Professor, Department of ECE, Hindusthan Institute of Technology. Coimbatore,Tamil Nadu,India[2]

Abstract- The disease is curable if detected early stage. Screening is carried out on the basis of mammogram; this is used in x-ray image to reveal lumps in the breast. Calcium deposit can also indicate the existence of a tumor in breast. Mammography is proven as efficient tool to detect breast cancer before clinical symptoms appears digital mammography is currently considered as standard procedure for breast cancer diagnosis, various artificial intelligence techniques are used for classification problems in the area of medical diagnosis. Several type of feature extraction is from digital mammograms including position feature, shape feature and texture features etc. Feature extraction of image is important in mammogram classification. These features are extracted by using image processing. Texture features have proven to be useful in differentiating normal and abnormal cells. Extracted texture features provide information about textural characteristics of the image. MLE (Maximum Likelihood Estimation) and wavelet transforms is used to calculate the area and also showing the affected area. This helps to determine the depth of tumor. Here '0' is showing as black and '1' is showing as white. Pre-processing method used as a small neighborhood of a pixel in an input image to get a new brightness value in output image, also called as Filtration. Breast cancer is a type of cancer originating from breast tissues, and most commonly this is originated from the inner lining of milk ducts. Breast cancer occurs in human and other mammals also.

Keyword-Digital mammograms, Maximum Likelihood Estimation, Wavelet transform.

I. INTRODUCTION

Brest cancer is the most frequent cancer in women the disease is curable if detected early enough. The disease is curable if detected in early stages. Screening method is carried out on the basis of mammograms, x-ray images are used to reveal lumps in the breast. The existence of tumor is occurred by calcium deposits. Digital mammography is currently as standard procedure for breast cancer diagnosis, various techniques are used for classification problem in the area of medical diagnosis. Feature extraction of image is important step in mammogram classification. These features are extracted using Digital Image Processing. MLE and DWT are used to calculate the area and also showing the cancer affected areas. Tumor cells as white in color .Breast cancer is a type of cancer originating from breast tissue , most commonly from inner lining of milk ducts or the lobules that supply the duct with milks.

Breast cancer occurs in human and other mammals. While the over whelming majority of human cases occur in human cases occur in women, male breast cancer also occur. The characteristics of the cancer determine the treatment, which may include surgery, medication (hormonal therapy and chemotherapy) radiation and immunotherapy. Prognosis and survival rates for breast cancer vary greatly depending on the cancer type, stages, treatment, and geographical location of the patient.

Cancer

Cancer medically as a malignant neoplasm, is a board group of disease involving unregulated cell growth. In cancer, cell divides and grows uncontrollably and which forming malignant tumors, invading nearby part of the body. The cancer can spread all part of the body through the lymphatic systems or blood streams. Not all tumors are cancerous; benign tumors do not invade neighboring tissues and do not spread throughout the body.

There are over 200 different known cancer that affected humans. The causes of cancer are complex, and only partially understood. Many are known to increase the risk of cancer, including dietary factors, tobacco use, dietary factors, exposures to radiation, and lack of obesity, physical activity, and environmental pollutants. These factors can damage genes or combine with existing genetic fault within cells to cause cancerous mutation.

Many cancers could be prevented by non-usage of smoking, eating vegetables fruits and whole grains in a large amount, less amount of meat and refined carbohydrates, helps to maintain a healthy weight, minimizing sunlight

ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2014

exposure, exercising and being vaccinated against some infectious diseases. Cancer can be detected in different ways, including the presence of certain sign and symptoms, screening test, or medical imaging.Cancer is primarily an environmental disease with 90-95% of cases attributed to environmental factors and 5-10% due to genetic. Common environmental factors include tobacco (25-30%), diet and obesity (30-35%) infections (15-20%) radiations, hereditary, are the reasons for causing cancer.

Risk factors of breast caner

The primary risk factors breast cancers are female sex and older age. Other potential risk factor include: lack of childbearing or lack of breastfeeding, higher levels of certain dietary patterns, and obesity. Most types of cancer are easy to diagnose by microscopic analysis of a sample or biopsy of the affected area of the breast. The two most commonly used in screening methods, physical examination is used in breasts tissues by a health care provider and mammography; this can offer an appropriate likelihood that a lump is cancer. And also detect some other lesions, such as a simple cyst. Many cancer is could be prevented by reduce smoking, eating more fruits, vegetables and whole grains, eating less meat and less carbohydrates, exercising, and being vaccinated against some disease.

Cancer can be detected in different number of ways. The presence of certain can be detected in a number of ways, including the presences of certain sign and symptoms, screening test or medical imaging.MLE helps to determine area of tumor affecting portions. And which also calculates the area of tumor affecting portions. It's a method of estimating parameters in a statistical model; the Maximum Likelihood Estimation provides estimates for the parameters, and Example is coin tossing. This helps to determine tumor and non-tumor cells. Regression model were apply to classify.

II. LITERATURE REVIEW

Pietro Perona, Jitendra Malik has proposed the scale-space techniques introduced by Wilkins involve generating coarser resolution image by convolving the original image with a Gaussian kernel. This approach has a major drawback: it is difficult to obtain accurately the locations of the "semantically meaningful" edges at coarse scales, and introduce a class of algorithms that realize it using a diffusion process. The algorithm involves elementary, local operations replicated over the image making parallel hardware implementations feasible.

Randovan and Torfinn Taxt had proposed a new method for two-dimensional deconvolution of medical ultrasonic image is presented. The spatial resolution of the deconvolved image is much higher compared to the common images of the fundamental and second harmonic. Broadband radio frequency image data are deconvolved instead of common fundamental harmonic data. The method was instead of common fundamental harmonic data. The method was validated on image data recorded from a tissue-mimicking phantom and on clinical image data.

Jialin Shen, Yuanyuan Wang, and Jianguo YU, Weiqi Wang proposed, The accurate boundary extraction is an essential preprocessing step for computerized analysis of a breast ultrasonic image Firstly, a rectangular region-of-interest (ROI) is manually selected from the ultrasonic image, followed by the ROI-based pre-processing for the noise reduction and image enhancement. Then an initial boundary of the tumor is obtained using the wavelet transform. Experiments on 45 breasts have shown that this proposed method.

Lu Zhang, Christine Cavaro-M'enard and Patrick Le Callet, proposed through several objective image quality assessment methods originally proposed for natural image and videos have been used in the context of medical images. This paper presents a review on some key issues that must be considered for the objective quality assessment of still radiographic images acquired from the acquisition system of varied imaging modalities.

III.METHOD

A. Database of Mammogram

The Digital Database for screening Mammography (DDSM) is a resource for use by the mammographic image analysis research in the development of the development of computer algorithms to aid in screening. Each study includes two image of each breast (age at time of study, ACR breast density rating for abnormalities, and image information (scanner, spatial resolution). Distribution includes normal, binomial, Poisson (count data) and exponential (positive continuous), Gaussian, binomial, gamma, inverse Gaussian, quasi binomial, quasi poison. Normal functions includes identity, log it, log, reciprocal, power, and which depends on distribution. Image containing suspicious areas have associated pixel-level "ground truth" information about the locations and types of suspicious regions. Also provided is software both for accessing the mammogram and truth images and for calculating performance for automated image analysis algorithms. Texture analysis of mammograms-The texture of a mammographic analyzed

ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2014

based on the difference between high and low gray levels in it. Various textures related parameters of mammographic image help us to determine them as normal or abnormal.

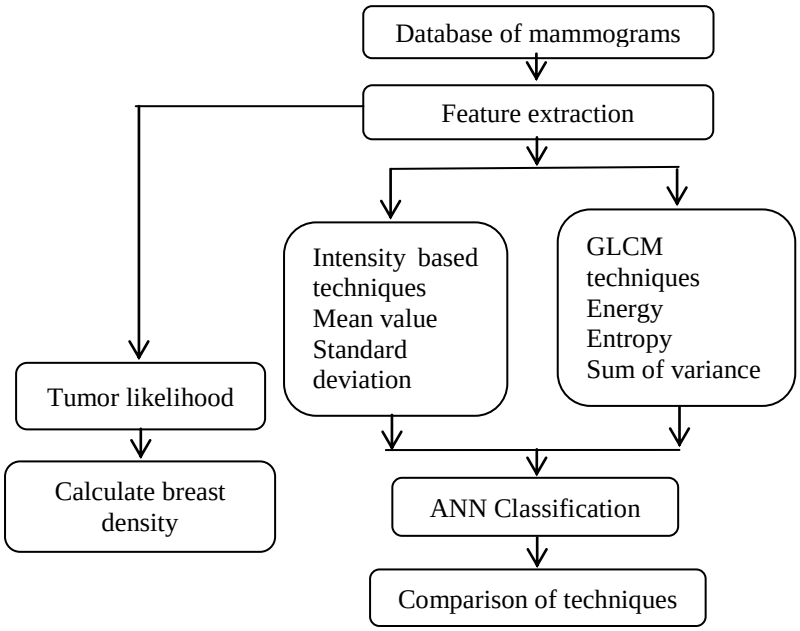


Fig. 1 Block diagram for differentiating tumor and normal cells

B. Maximum Likelihood Estimation (MLE)

Maximum Likelihood Estimation (MLE) is a method of estimating the parameters of a statistical model. When applied to set data and given a statistical model, maximum likelihood estimation which provides estimates for the model's parameters, example is coin tossing. Its help to distinguish between tumor and non-tumor cells, a regression model were applied to classify.

A MLE is a parameter estimate that is most consistent with the sampled data. It maximizes the likelihood functions. The method of maximum likelihood selects the set of values of the model parameters that maximizes the likelihood function. The Maximum likelihood Estimation gives a unified approach to estimation; it's defined in the case of the normal distribution and many other problems.

The principle of maximum likelihood estimation (MLE), originally developed by R.A.Fisher in the 1920s, states that the desired probability distribution is the one that makes the observed data "most likely" which means that one must seek the value of the parameters vector that maximizes the likelihood functions. Distribution steps are identify the distribution, then choose analysis then loss function. Parameter value that maximizes likelihood and model comparison in MLE. Loss function which uses $(y\text{-}y^2)^2$ as the loss function and tries to minimizes the sum of this quantity (across rows).

MLE loss functions depends on the assumed distribution, Negative log likelihood becomes our loss functions. The maximum likelihood selects the set of values of the model parameters that maximizes the likelihood functions. Maximum likelihood estimation gives a unified approach to estimate, which is well defined in the case of the normal distribution and any other problems.Best parameters are obtained by maximizing the probability of obtaining the samples.

Bayesian methods view the parameters as random variables having some known prior distribution. Has good coverage properties as the sample size increases; estimated parameter value approaches the true value as increases. Parameters describe the characteristics of a population. Their values are estimated from some samples, collected from that population. Algorithm for Maximum Likelihood Estimation is, most statistical method is designed to minimize error.

ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2014

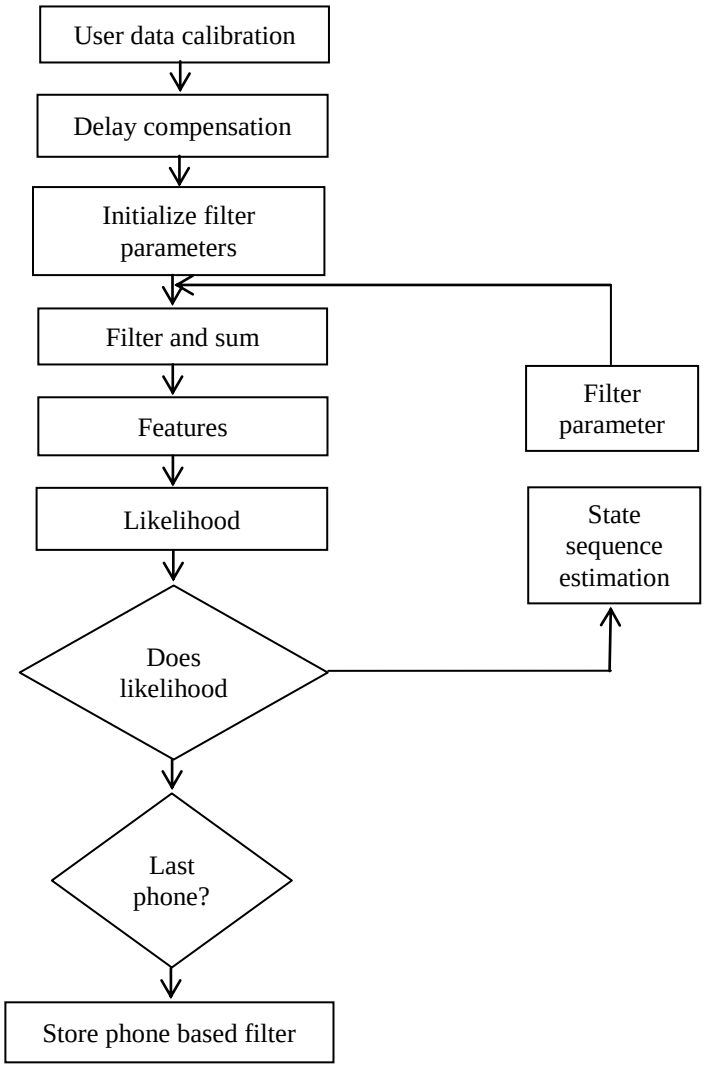


Fig. 2 Flowchart for MLE

C. Discrete Wavelet Transform (DWT)

A DWT is a wavelet transform for which the wavelet transform decomposes a signal into a set of basic functions are called wavelet. One use of wavelet approximation is in data compression. Wavelet transform decomposes a signal into set of basic functions. These basic functions are called wavelet. Wavelet are obtained from a single prototype wavelet y(t) called mother wavelet by dilatation and shifting. Discrete Wavelet Transform which transform a discrete time signal to a discrete wavelet representation. It convert an input series x_0 , $x_{1,\ldots\ldots}x_m$, into one high-passwavelet coefficient series and low-pass wavelet coefficient series (of length n/2 each). Good for signal having high frequency resolution and poor time resolution at low frequency components for long duration.

THEORY OF WT, the wavelet transform is computed separately for different segments of the domain signal at different frequencies. Multi-resolution analysis, analyzes signal at a different frequencies giving different resolutions.

1D DWT

Separate the high and low-frequency portions of a signal through the use of filter, one level of transform, signal is passed through G& H.

ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2014

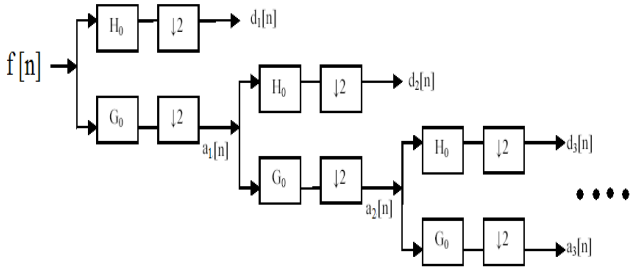


Fig.3 Multiple levels (scales) are made by repeating the filtering and decimation process on low pass outputs.

2D DWT for image

To use wavelet transform for image processing we must implement a 2D version of the analysis filter bank is first applied to the columns of the image and then applied to the rows. If the given image (figure 4), N1 rows and N2 columns. The 2D synthesis filter bank combines the four sub band images to obtain the original image size of N1 & N2.

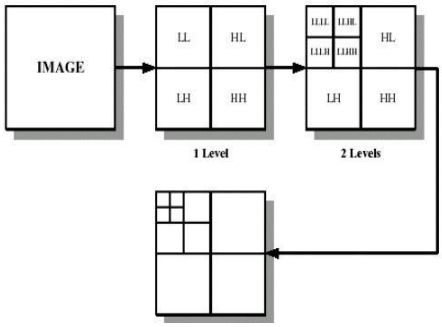


Fig.4 Subband image

Mother wavelet, the efficiency reason, one prefers continuously differentiable functions with compact support as mother (prototype) wavelet (functions). The wavelet transform is often compared with the Fourier transform can be viewed as a special case of the continuous wavelet transform with the choice of the mother wavelet. The main difference in general is the wavelets are localized in both time and frequency where as the standard Fourier transform is only localized in frequency.

The short time Fourier transform (STFT) is more similar to the wavelet transform, in that it is time and frequency localized, but there is issues with the frequency or time resolution trade off. Wavelet gives a better signal representation using multi-resolution analysis, with balance resolution at any time and frequency. The discrete wavelet transform is computationally less complex, O(N) time as compared to O(NlogN) for the fast Fourier transform.

This computational advantage is not inherent to the transform, but reflected a choice of the logarithmic division of frequency and in contrast to the equally spaced frequency division of the FFT which uses the name basic functions as DFT (Discrete Fourier Transform). It is also important to note that this complexity only applies when the filter size has no relation to the signal size. A discrete wavelet without compact support such as the Shannon wavelet would require $O(N^2)$.

HAAR WAVELET

The Haar wavelet is a sequence of rescaled "square-shaped" functions which together form a wavelet a discrete wavelet transform (DWT) is any wavelet transform for which the wavelet transform for which the wavelets are discretely sampled family or basics. A wavelet analysis is similar to a Fourier analysis is similar to Fourier over an interval to be represented in terms of an orthogonal function basics.

The Haar sequence was proposed in 1990 by Alfred Haar. The study of wavelet, and even the term 'wavelet', did not come until much later. As a special case of a Daubechies wavelet, the Haar wavelet is also the simplest possible wavelet. The disadvantage of the Haar wavelet in technical is that is not continuous and therefore not differtiable.

D. BREAST DENSITY

Area of breast is calculated by MLE and DWT. These also helps to showing the portions which affected by tumors. Area is calculated by Pixels

ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2014

PIXEL

A pixel is generally thought of as the smallest single components of a digital image, its definition is highly context-sensitive. This is depending on context; there are several terms that are synonymous in a particular context, such as pel, sample, bytes, bit, dot, spot etc. the term "pixel" can be used in the abstract, or as s unit of measures of resolution.

The measure dots per inch (dpi) and pixels per inch (ppi) are sometimes used interchangeably, but this have distinct meanings, especially for printer device, where dpi is a measure of a printer density of dot placement. In a digital imaging, a pixel or pel (picture element) is a physical point in a raster image, or this is the smallest addressable elements in a display device.

PRE-PROCESSING

Pre-processing images commonly involves removing low frequency from background noise; this normalizing the intensity of the individual particle images. Digital Image preprocessing is the techniques of enhancing data images prior to computational processing. The output is said to been preprocessed from the input data. Preprocessing methods use a small neighborhood of a pixel to get a new brightness value in output image. Such preprocessing operation is called Filtration. Parameters describe the characteristics of population. Their values are estimated from samples collected from those populations.

E. INTENSITY BASED FEATURES

Intensity based features are first order statistics depends only on individual pixel values. The pixel intensities are Mammogram proposed method simplest available feature useful for pattern recognition. The intensity and its variation inside the mammograms can be measured by features like mean and standard deviation.

Mean value

The mean value gives the average intensity value of an image. Mammographic image that contain micro calcification have a higher mean than those of normal images. Where mean calculated from the given image. Here 'i' indicates the row of the image, 'j' indicates the column of the images, and P (i, j) is the cell denoted by the row and the column of the image.

$$\mu = {}^{1}/_{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} p(i,j)$$

Standard deviation

The standard deviation is a parameter closely associated with the mean. These refer to the dispersion of values in a digital mammographic image around the mean value. The SD is;

$$\text{SD}=\sqrt{(mean)2}$$

F. GLCM Features

The gray level co-occurrence matrix (GLCM) texture measurement is a method to analyze image texture. It is robust method that has been developed for calculating first and second order texture features for image. The GLCM matrix is a tabulation of how often different combinations of gray level occur in an image.

Energy

Energy represented the orderliness of a mammographic image. Energy is generally given by the mean squared of a mammographic image.

$$\text{E}= \sum_{j\ i=}^{n-1} P\ (i,j)2$$

Entropy

The amount of disorder in a mammographic image is called as entropy. The entropy value is high in micro calcification. This is because the variation in intensity values in the image is high due to the presence of white calcification spots.

$$\text{H}= \sum_{i=0}^{n} p(i.j) \log P(i,j)$$

IV.RESULT

Differentiating normal cells and tumor cells and these also help to calculate the area of tumor affecting portions. It shows the area of calculation in unit of pixel Maximum Likelihood Estimation (MLE) and Discrete Wavelet Transform help to determine the area of tumor. Intensity based techniques and GLCM techniques are used. Intensity based

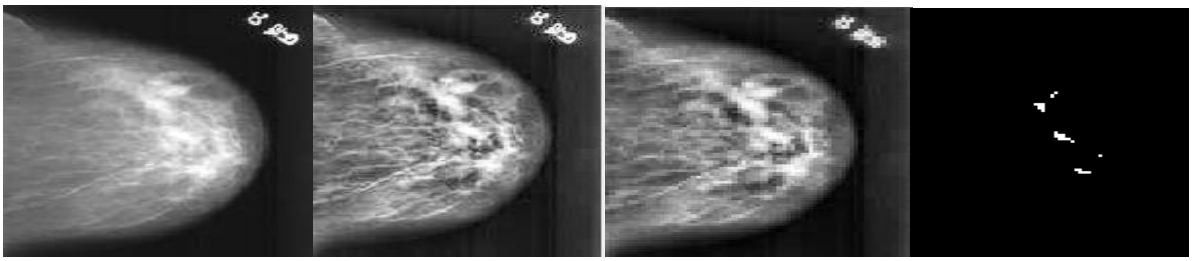
ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2014

techniques helps to determine mean value, standard deviation. And GLCM techniques are helps to determine energy, entropy, contrast, correlation, homogeneity and sleekness.



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

Fig.5 (a) Input image; (b) Pre-processing image; (c) Wavelet image; (d) Image differentiating tumor and non-tumor cells.

Table 1.Textual Features of Breast Cancer

Contrast	0.0539	0.0607	0.04099
Homogeneity	0.9564	0.9866	0.9895
Correlation	0.993	0.9947	0.9875
Energy	0.5429	0.480	0.5963
SD	71.196	82.563	45.87
Mean	41.320	56.4533	26.147
Entropy	0.92661	0.9999	0.967
Skewness	1.338	0.9773	1.6151
Type	Normal	Cancer	Normal

V. DISCUSSION

Intensity based and GLCM techniques are used here, calculate the energy, entropy, contrast, correlation, homogeneity, and sleekness. For GLCM contrast is 0.048442, homogeneity is 0.986861, correlation is 0.990888, energy is 0.555255, entropy is 0.999986, skewness is 1.41176. For intensity based techniques mean is 34.3779 and standard deviation is 57.4773.

(a) (b)

Fig. 6 (a) Cancerous image; (b) Normal cell

VI. CONCLUSION

Digital mammography is currently as standard procedure for breast cancer diagnosis, various techniques are used for classification problem in the area of medical diagnosis. Feature extraction of image is important step in

ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2014

mammogram classification. These features are extracted using image processing techniques. Area of tumor is calculated by the MLE (Maximum Likelihood Estimation) and DWT (Discrete Wavelet Transform) tumor affecting portions are denoted. The normal and cancerous cell is showing separately, and entropy, mean, standard deviation, energy, skewedness' etc. are calculating d from the database image. By using MATLAB software, its applicable in biomedical. Low sensitivity, less tolerance of MSE, improved PSNR and more accuracy for area calculation are its advantages.

REFERENCES

[1] A.Jemal, F.Bray, M.M. Canter,J. Feraly, E. Ward, and D.Forman, "Global cancer statistic," A A Cancer journal for Clinicians., 2011.
[2] A.Malich, D. R. Fischer, and J. B ottcher, "CAD for mammography: the technique, results, current role and further developments," European Radiology, 2006.
[3] D. P. Chakaraborty, "Validation and statistical power comparision of method for analyzing free-response observer performance studies," Acad. Radiol., vol6,pp, 65-70, 1979.
[4] F. J. Gilbert, S. M. Astely, M. G. C. Gillan, O.F. Agbaje, M. G. Wallis, J. James, C. R. M. Boggis, S.W. Duffy, and CADETII Group, 1675-1684 "Single reading with computer-aided detection for screening mammography," New England Journal of Medicine, vol. 359, pp.
[5] K. Kelley, J.Dean, W. Comulada, and S. Lee, "Breast cancer detection using automated whole ultrasound and mammography in radio graphically dense breasts," European Radiology, vol. 20, pp.734-742, 2010.
[6] M.B. Dillencourt. H. Samet, and M. Tamminen, "A general approach to connected-component labeling for arbitrary image representation",
[7] M. J. Morton, D. H. Whaley, K. R. Brandt, and K. K Amrani, "Screening mammograms: interpretation with computer aided detection-prospective evaluation," Radiology, vol 19, p. 1425, 2006.
[8] N. F. Boyd, H. Guo, L. J. Martin, L. Sun, J. Stone, E. Fishell, R. Q. Jong, G. Hislop, A. Chiarelli, S. Minkin, and M. J. Yaffe,2007 "Mammographic density and the risk and detection of breast cancer, " New England Journal of Medicine, vol 356, pp.227-236.
[9] R.A. Groeneveld and G. Meeden, "Measuring Skewness and kurtosis," J.R. Stat. Soc.,vol. 33, pp. 391-399, Dec. 1984.
[10] R.R. Picard and R.D. Cook, "Cross-validation of regression models," J. Am. Sat. Assoc., vol, 79,pp. 575-583, Sep 1984.
[11] S. Holm, "A Simple sequentially rejective multiple test procedure, " Scand. J. Stat., vol. 6, pp.65-70, 1979.
[12] S. M. Astely and F. J. Gilbert, "Computer-aided detection in mammography." Clinical Radiology, vol. 59, pp. 390-399, 2004.

BIOGRAPHY

ANU ALIAS

She was born in Perumbavoor, Ernakulam in Kerala, India. She completed her B.E. Eectronics & Communication Engineering from CMS College of Engineering, ANNA University Chennai in 2012. She is currently doing the Post-graduation in VLSI Design & Embedded Systems from Hindusthan Institute of Technology,Coimbatore, Tamil Nadu, now working on the research project in Digital Image Processing.

B. PAULCHAMY

Hewas born on 06th February 1975. He is currently working as a Assistant Professor and Head of ECE department in Hindusthan Institute of Technology, Coimbatore, India. He is doing his (Ph.D), and his area of specialisation in Signal Processing. And he is also a member in ISTE, he have 12 year of professional experience. He published 10 International journal and 5- International, 27-National Conferences. HE also published a book in Digital Image Processing.