Breast Cancer-Early Detection and Classification Techniques: A Survey

Anu Appukuttan
Dept. of Computer Science
College Of Engineering Poonjar

Sindhu L.
Dept. of Computer Science
College Of Engineering Poonjar

ABSTRACT
Breast Cancer is the most common incursive cancer which is found in females all through the world. Of all the female cancers it comprises of 16% and it accounts for 22.9% of invasive cancer in women. of all the cancer deaths 18.2% are from breast cancer which includes males and females.. As the modern science is improving many researches and techniques have been emerged to eradicate this dreadful disease. So there is a need of an automated computer aided diagnosis system and it is proposed here. This survey paper focus on highlighting different techniques on enhancement, detection and classification of breast cancer along with its accuracy.

General Terms
Classification, Database.

Keywords
Accuracy, Breast Cancer, CAD, Classifiers, Detection, Enhancement. MIAS.

1. INTRODUCTION
Cancer is the most dreadful disease which is a collection of related diseases. In every cancer types, some of the body's cells begin to divide involuntarily and spread in to surrounding tissue.

Likewise, breast cancer arises from breast cell tissues. Breast, like any part of the body, consists of billions of microscopic cells. These cells start to multiply uncontrollably causing breast cancer. There are mainly 2 types of breast cancer, a ductal carcinoma is the most common cancer where cancer begins in the milk duct and the cancer that begins in the lobules are called lobular carcinoma[1]. At the ancient times, the diagnosis and treatment proves to be fatal without efficient techniques. At each stage, the vigorosity of this disease and the death rate is increasing. So there is a need of early detection of breast cancer to reduce the death rate. So automated computer aided detection is inaveitable.

There are no causes of breast cancer, we can point it as only risk factors. It may be genetic or environmental. Genetic factors include family history, personal health history, menstrual and reproductive history, dense breast tissue, certain genome changes, age, gender etc. The environmental factors include obesity, poor diet, alcohol consumption, radiation, lack of physical activity etc [2].

The initial symptom of a breast cancer is the formation of a lump. This is due to tiny deposits of calcium called microcalcifications and tumors called circumscribed masses. These tumors are generally Benign and malignant. The benign tumors are generally non aggressive and non cancerous. They will not spread to other body parts. The malignant tumors are considered as cancerous and aggressive as well as spread all over the body cells i.e, metastatic in nature [3]. Using a CAD system ,the classification whether a tumor is benign, malignant, normal is done through different techniques like neural networks, fuzzy c means clustering, SVM classifiers etc.

Different imaging techniques are possible for early detection of breast cancer. These include X ray imaging, digital mammography, ultrasound imaging, MRI screening etc. The digital mammography is widely used nowadays due to its advantages over others. X ray imaging is usually used for finding the signs of the cancer. While using a mammogram generally investigate the problem. Mammogram uses X rays to create images of the breast. Earlier there are film mammographies in which the images are stored on films, but now digital mammograms are widely used since they are stored directly on computers as well as all the nook and corners will be visible for easy detection. In breast ultrasound the images are created using sound waves. But this is not used nowadays as it is done with a handheld device, it will generate false positives and false negatives when the person who operates it is not well experienced or skilled and thus the quality of the image will vary.

The detailed detection of the images is done using feature extraction and texture extraction techniques. This itself opens another research area as a wide variety of techniques have been used for segmentation, feature extraction, enhancement. Done mainly by wavelet techniques, clustering, using GLCM matrix etc which are described clearly in the related works. So the ultimate aim of this survey is to provide different enhancement, detection and classification techniques for early breast cancer detection.

2. ENHANCEMENT METHODS
Image enhancement means making the processed image more useful for analysis. Enhancement is achieved by intensifying the contrast and edges and by suppressing the noise and background areas. The need for image enhancement arises because of its subtle and complex nature. The studies have shown that even the experienced radiologists can do errors while detecting breast cancer in radiographic findings. Thus the computer aided diagonosis demands an enhancement in mammographic images thereby increasing the efficiency in detection based on advanced image processing and pattern recognition techniques. So the radiologists can have a second opinion in the diagonosis.

2.1 Fractal Modeling
Fractal is considered as a mathematical set that exhibits a repeating pattern at every scale. It has been used to enhance the microcalcifications in mammographic images. . The basic property of fractal objects is the local self similarity which is possessed by mammographic structures. So the fractal model will construct these tissue patterns and are taken out from the

original image, in this way microcalcifications will be enhanced.

2.2 Wavelet Transform

The wavelet transforms include morphological operation, tophat operation, dyadic wavelet transforms. The advantage of using wavelet transform is that it has good time frequency and localization property and thereby increasing the contrast of images. The top hat operation states that one can extract various image contexts by taking the difference between one processed by closing and opening operation with a specified structural element. The dyadic wavelet transform preserve the subband images remain same without duplication.

2.3 Region description method

It is a simple image segmentation method. The unwanted noise can be removed using this method along with a thresholding technique. To suppress the pectoral muscle `this method measures the properties of image region such as area, perimeter, shape, compactness etc.

3. DETECTION AND CLASSIFICATION METHODS

Detection and classification is done with a wide variety of techniques. Some of the techniques are below

3.1 Detection Techniques

3.1.1 Wavelet Neural Network

The wavelets has got both spatial and domain characteristics and thus can be used to find out the abnormalities. The ANN is the best tool for the diagnosing the pathological images especially of cancers and precancers. By combining the wavelet theory and neural network we will get best detection method for breast cancer.

3.1.2 Curvelet transforms

It is a new image representation. The main feature is that it has efficient image coding, geometrical features, optimal object representation with edges, image reconstruction, optimal sparse representation of wave propagators than wavelets. Here feature vector is taken as curvelet transform coefficients.

.

3.1.3 Contourlet transform

The minute geometrical features of the images such as smooth contours are detected using contourlet transform. This has got multiscale and directional properties. Through this a mammographic image is decomposed in to various directional subband at different scales using 2D filter bank.

3.1.4 GLDM

For an image GLDM calculates the GLDM probability density functions. Commonly used for extracting statistical textural features of a mammographic image. Five texture features are defined: Contrast, Angular Second moment, Entropy, Mean and Inverse Difference Moment from each density functions.

3.1.5 Gabor Filter and Histogram Equalisation

Image quality is increased using histogram equalization method. Gabor filter used to remove the noisy signals present in the image. A 2-D Gabor function is a Gaussian modulated by a sinusoid. It is a non-orthogonal wavelet.

3.2 Classification techniques

3.2.1 PSOWNN

PSO has been used for optimization problems. The optimal features of the images are selected and then classify the images based on wavelet neural network.

3.2.2 SVM Classifier

It is a machine learning method which uses a hyperplane that maximizes the margin in the training data to classify binary classes. Support vectors are the training data along the hyperplane. The distance between the support vectors and the class boundary is the margin. The decision planes that define decision boundaries are the basic idea of SVM.

3.2.3 K Means Clustering

The basic idea of k means clustering is that the n dataset is partitioned in to k clusters, in which each extracted image belongs to the cluster with nearest mean i.e, mean is taken as the criteria. Likewise all the images are classified.

3.2.4 SRAN Classifier

In the SRAN system, the training sample record arrives one by one and the network adapts its parameters on the basis of the difference in knowledge between the network and the current sample record. Uses basic concept of RBFN.

3.2.5 Probabilistic Neural Network

PNN is a form of Probabilistic density function, training is very easy and fast for PNN. The computational load in the training phase is transferred to the evaluation phase in PNN which makes it different from other classifiers.

4. RELATED WORKS

Here we are discussing different techniques used in enhancement of mammographic images, detection and classification of breast cancer.

4.1 Analysis Of Enhancement Techniques Of Mammographic Images.

Huai Li et.al[4] proposed a deterministic fractal approach to enhance the microcalcifications and to model the parenchymal and ductal patterns. The affine transformation parameters are used for modeling this. The microcalcifications are having different shapes, the major task here was to differentiate or extract the microcalcifications from background images The enhancement is evaluated by decomposing the mammogram in to 4 levels. The filtered version of the image is reconstructed by partially selecting the subimages in the second and third levels. The low frequency background structure was removed and the microcalcifications were enhanced after the reconstruction.

The enhancement of microcalcifications were evaluated with other works and compared with CII, the noise level, the PSNR and the ASNR and found out that the noise level of fractal approach was much lower. One issue is that fractal modeling approach needs enormous large computation in algorithm implementation.

Vijayakumar Guturu et.al[5], using wavelet transform enhanced the contrast of mammographic images. For this a wavelet based level dependent thresholding algorithm and a modified mathematical morphology algorithm was used. Simple operations like erosion and dilation are combined as mathematical morphology and are done with tophat operation. This reduces noise. As a wavelet transform technique dyadic

wavelet transform has been used. Evaluated using CII and studies with Visushrink and Normalshrink and showed greater accuracy.

The paper proposed by B.Senthilkumar et.al.[7], modified LRM(Local Range Modification) for enhancement and noise removal and proposed MLRM. Fuzzy C Means Clustering is used for cluster detection and enhancement and for the segmentation and enhancement of these clusters a tree structured non linear filter is used. MLRM differ from LRM in 2 ways. It computed 48*48 pixels in the first pass instead of 51*51 in LRM and used 8 grid points instead of 4 in LRM. This arised complexity in computation, but the image quality is good and satisfactory w.r.t LRM.

Ojo.J.A.et.al [6], proposed a preprocessing method for extraction of pectoral muscle and removal of artefacts in mammogram. The unwanted noise removal and the extraction of pectoral muscle is done with region description method which includes splitting and merging. The need for removal of pectoral muscle arises because it is the dense region in the breast which resembles the abnormality of breast cancer, so there is a greater chance of false detection. The extracted region is then filtered using a global thresholding technique. Suppression was done by polygon approximation method. MIAS database images have been used.

.

TABLE 1: Comparison of different enhancement techniques

Author	Year of publication	Enhancement technique	Advantages/Disadvantages
Huai Li et.al[4]	1997	Fractal modeling scheme	Huge computation is needed.
Vijaya Kumar Gunturu et.al[5]	2010	Morphological Top-hat, Dyadic wavelet-based-level dependent thresholding algorithm.	Preserve overall shape of spots.
Ojo J. A et.al[6]	2014	Region description , Polygon approximation	Time consuming
B.Senthilkumar et.al[7]	2013	Modified LRM and Fuzzy C means Clustering	Complexity in computation Good Image quality

4.2 Analysis of detection and classification techniques of mammographic images

Here the different techniques for detection and classification are discussed and compared.

J. Dheeba et.al[8] proposed a Particle Swarm Optimised Wavelet Neural Network (PSOWNN) to classify and detect breast cancer. WNN posses both wavelet and neural network properties, here the ROI detection is done through Global thresholding technique.WNN along with PSO proves to be best for classification as it decreased false negatives and false positives. From the mammographic images the laws texture energy measures are extracted with an abnormality detection algorithm. The textural features are extracted through a windowing operation and by convolution kernels applied to ROI.216 mammographic images obtained from mammographic screening centres have been used. Advantage is that this method increased convergence of back propagation algorithm error and the disturbances in learning are avoided.

K. Subashini et. al[11] proposed breast cancer detection through ultrasound images. Wavelet domain techniques i.e, DWT translates the images in to wavelet coefficients to remove noise. Here the noise representing coefficients are suppressed and image features are enhanced. Segmentation is done with active contour model and the texture features are extracted using auto covariance coefficients. The back propagation neural network are used for classification which found better in performance than linear classifiers.

Ibrahima Faye et.al[9] proposed multiscale curvelet transform to detect breast cancer. Curvelet transforms are having different scale levels, a set of biggest coefficients are extracted from 2,3,4,5,6,7 scale levels. These coefficients are feature vectors used for classifying the mammograms. Feature extraction is done by transforming ROI's n to different scale levels. A supervised classifier is constructed using Euclidean distance. The distance between feature vector and core vector (mean of image set) is calculated for classification. MIAS database is used. Accuracy at different scale levels are computed and found that the scale 2(98.59%) is having more accuracy than scale 5, this is because as the scale increases the redundant data also increases.

Y. Ireaneus Anna Rajani et.al[12] aims at finding solutions to detect the doubtful regions in poor contrast images. This method was tested on 75 images of MIAS database. For the enhancement filtering,, tophat, DWT and for segmentation thresholding is used. The features like area, centroid, eccentricity etc are extracted. The classification is done using SVM and provide an accuracy of 88.75%.

S. Julian Savari et.al[10] proposed a paper that uses histogram equalisation to increase the image quality. The intensity features are extracted and computed to calculate volumetric values. The classification is done by K means clustering algorithm. Noise is removed by Gabor filter. The dataset used are MIAS and DDSM database. Provide 99% classification accuracy.

Neetha Jog et.al[14] used GLDM and Gabor features as extraction methods with SVM and KNN classifiers. Here MIAS database is used. Gabor feature along with wavelets are also used. Basic idea of SVM was to find an hyperplane to separate D dimensional data in to two classes finely. KNN classify objects based on nearest training samples in the feature space. Result showed that SVM+GLDM provide 95.83% accuracy than KNN+GLDM and SVM+Gabor filter has 71.83% accuracy than KNN+Gabor Filter.

S. Deepa[15] et.al discussed the different ranges of Contourlet Coefficient Cooccurance Matrix features in the analysis of mammogram images and classification. The ROI is enhanced using histogram equalisation and are reduced using contourlet transform. The cooccurance matrices are produced for various directions. Feature selection is done by Sequential Floating forward Selection Algorithm(SFFS). Classification is done through Probabilistic Neural network (PNN). The demerit is that the system is not automated. MIAS database is used providing 92.5% accuracy.

S. Shanthi et.al[13] proposed a CAD system. Initially all the noise has been removed using morphological operations.ROI is identified automatically using Fuzzy C means Clustering. Histogram features ,Gray level Concurrence features and wavelet energy features are used for designing feature database. Classification is done by SRAN(Self Adaptive Resource Allocation Network). From the training set those samples that are not advantageous are deleted, but lack of optimal feature subset is a demerit.

TABLE 2: Comparison of different approaches in detection and classification

Author	**Publication Year**	**Detection method**	**Classification method**	**Accuracy**	**Database**	**Advantages/Disadvantages**
J.Dheeba Albert Singh S.Tamil Selvi[8]	2014	Wavelet Neural Network	Particle Swarm Optimised Neural Network (PSOWNN)	93.67%	216 mammographic images from mammographic centers	Difficulty in learning is avoided. Convergence of back propagation algorithm error is Increased.
Mohamed Meselhy Eltoukhy, Ibrahima Faye1, Brahim Belhaouari Samir[9]	2010	Curvelet transforms with biggest coefficients(feature vectors)	Euclidean distance	98.59%	MIAS db	Stable and Efficient. Cause redundancy on increasing no: of scales.
S. Julian Savari Antony, Dr.S.Ravi[10]	2015	Gabor Filter(feature extraction), Histogram equalization(enhancement)	K-Means Clustering Algorithm	99%(Classification accuracy)	MIAS db, DDSM db	Reduces space and time complexity
K.Subashini, K.Jeyanthi[11]	2014	Wavelet Domain techniques, Wavelet Shrinkage.	Back Propagation Neural network		Ultrasound Images	generate false positives and false negatives when operator is not skilled.
Y.Ireaneus Anna Rejani, Dr.S.Thamarai Selvi[12]	2009	Filtering,tophat ,DWT(enhancement) Thresholding (feature extraction)	SVM Classifier	88.75%	MIAS db(75 images)	Fast and easy.
S.Shanthi, V.Murali Bhaskaran[13]	2012	Fuzzy-C-Means(ROI selection), Histogram, Greylevel Concurrence, wavelet features (Feature db)	SRAN(Self Adaptive Resource Allocation Network)	Precision and recall calculated and found better than other methods	MIAS db	Deleted samples that are not advantageous from training set. Disadvantage is that optimal feature subset is not there

Neeta Jog, Arvind Pandey[14]	2015	GLDM, Gabor Feature	SVM, KNN	GLDM+SVM=95.83% SVM+Gabor Filter=71.83%	MIAS db	Found GLDM descriptor and SVM have more accuracy than KNN and Gabor Filter.
S.Deepa, Dr.V.Subbiah Bharathi[15]	2013	contourlet transform, Sequential Floating Forward Selection(Feature Selection)	Probabilistic Neural Network(PNN)	92.5%	MIAS db	Training is fast and easy. Not Fully automated System.

5. CONCLUSION

This survey paper concludes that there are several techniques that deals with pre-processing, enhancement, segmentation, feature extraction and classification of diagnosing images that gave different accuracies. Also found that the mammographic images are giving better accuracy than ultrasound images, MRI images etc and most of the works used MIAS database which contain 322 mammographic images. But in future ultrasound images can be used as it is devoid of radiations. Of these a reliable and effective method has to be found out. The advancement of curvelets are a promising approach as it has got many advantages than wavelets. The use of SVM also increases the accuracy than any other classifiers[12].

6. REFERENCES

[1] Every Women Counts, Resource for Health Professionals.

[2] National Breast Cancer accounts.

[3] A Review On Breast Abnormality Segmentation And Classification Techniques

[4] 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.*

[5] Vijaya Kumar Gunturu, Ambalika Sharma 'Contrast Enhancement of Mammographic Images Using Wavelet Transform',©2010 IEEE.

[6] Ojo J. A., Adepoju T. M., Omdiora E. O., Olabiyisi O. S. and Bello O. T, 'Pre-Processing Method for Extraction of Pectoral Muscle and Removal of Artefacts in Mammogram,' IOSR Journal of Computer Engineering (IOSR-JCE) *e-Volume 16, Issue 3, Ver. V (May-Jun. 2014).*

[7] B. Senthilkumar and G.Umamaheswari, 'Combination of Novel Enhancement Technique and Fuzzy C Means ClusteringTechnique in Breast Cancer Detection' Biomedical Research 2013*; 24 (2); 252-256.*

[8] J. Dheeba, N. Albert Singh, S. Tamil Selvi 'Computer-aided detection of breast cancer on mammograms: A swarm intelligence optimized wavelet neural network approach', Journal of Biomedical Informatics 49 (2014) 45–52, 2014 Elsevier Inc.

[9] Mohamed Meselhy Eltoukhy, Ibrahima Faye1, Brahim Belhaouari Samir, 'Breast cancer diagnosis in digital mammogram using multiscale curvelet transform', Computerized Medical Imaging and Graphics *34 (2010) 269–276.*

[10] S.Julian Savari Antony ,Dr.S.Ravi, 'A New Approach to Determine the Classification of Mammographic Image Using K-Means Clustering Algorithm', International Journal of Advancements in Research & Technology, *Volume 4, Issue 2, February -2015* .

[11] K.Subashini, K.Jeyanthi, 'Masses detection and classification in ultrasound images', IOSR Journal of Pharmacy and Biological Sciences (IOSR-JPBS),*Volume 9, Issue 3 Ver. II (May -Jun. 2014), PP 48-51*

[12] Y.Ireaneus Anna Rejani, Dr.S.Thamarai Selvi Noorul 'Early Detection Of Breast Cancer Using SVM Classifier Technique', International Journal on Computer Science and Engineering *Vol.1(3), 2009, 127-130*

[13] S. Shanthi, and V. Murali Bhaskaran ,'Computer Aided System for Detection and Classification of Breast Cancer', International Journal of Information Technology, Control and Automation (IJITCA) *Vol.2, No.4, October 2012*

[14] Neeta Jog, Arvind Pandey, 'Implementation of Segmentation and Classification Techniques for Mammogram Images', IOSR Journal of Engineering (IOSRJEN), *Vol. 05, Issue 02 (February. 2015*)

[15] S. Deepa, Dr.V.Subbiah Bharathi, 'Textural Feature Extraction and Classification of Mammogram Images using CCCM and PNN', IOSR Journal of Computer Engineering (IOSR-JCE) ,*Volume 10, Issue 6 (May. - Jun. 2013)*