OPEN ACCESS INTERNATIONAL JOURNAL OF SCIENCE & ENGINEERING

AUTOMATIC BREAST SEGMENTATION AND CANCER DETECTION USING SVM

R.S.Thakare[1], Dr. S. M. Deshmukh[2], Prof. V. R. Raut[3]

Instrumentation Engg. Government Polytechnic Yavatmal, India[1]
Electronics & Telecomm. Engg. P.R.M.I.T. & R , Badnera, Amaravati, India[2,3]
e-mail: aryadgire@gmail.com, hod_extc@mitra.ac.in, rautvivek@rediffmail.com

---

Abstract- Screening and early detection of breast cancer needs an automated system that identifies the breast cancer in the mammograms as early as possible. Breast Cancer is the most often identified cancer among women and major reason for increasing mortality rate among women. As the diagnosis of this disease manually takes long hours and the lesser availability of systems, there is a need to develop the automatic diagnosis system for early detection of cancer. An automated system that segments the mammogram masses and identifies the defect in the mammograms is proposed. The mammogram images are pre-processed by using median filter and adaptive histogram equalization. From the mammogram images features are extracted using Gabor algorithm and also by calculating mean and standard deviation of the image. The selected features were then classified using Support Vector Machine classifier. The classifier first identifies whether the input image is normal or abnormal. If the image is identified to be abnormal means the breast masses are segmented from the preprocessed images using Likelihood binarization algorithm. The segmentation algorithm segments the breast masses from the image based on the clustered result of group of pixels in the image. Then the mammogram image is classified into Benign or Malignant based on separate label and the features extracted. Finally the performance of the classifier is measured by calculating accuracy, sensitivity and specificity.

Keywords— *Breast cancer, Mammograms, Median filter, Gabor algorithm, SVM, Likelihood binarization algorithm*

I INTRODUCTION

Breast Cancer is the prime reason for demise of women. It is the second dangerous cancer after lung cancer**.** In the year 2018 according to the statistics provided by World Cancer Research Fund it is estimated that over 2 million new cases were recorded out of which 626,679 deaths were approximated. Of all the cancers, breast cancer constitutes of 11.6% in new cancer cases and come up with 24.2% of cancers among women**.** In case of any sign or symptom, usually people visit doctor immediately, who may refer to an oncologist, if required. The oncologist can diagnose breast cancer by: Undertaking thorough medical history, Physical examination of both the breasts and also check for swelling or hardening of any lymph nodes in the armpit [1]. Early diagnosis of breast cancer via mammographic screening is the current approach to reduce breast cancer mortality Interpreting mammograms, however, is a difficult task that requires special training and experience due to low prevalence of cancer in a screening population and superimposition of breast tissues in mammograms. Computer- aided diagnosis (CAD) offers a means to improve the efficiency of mammography and help radiologists to achieve higher diagnostic accuracy [3].

Mammography is one of the most reliable and effective methods for detecting breast cancer at its early stages. In developed countries, population-based mammography screening programs have been implemented. Women are encouraged to participate in regular breast examinations through mammography. In the U.S., annual mammographic screening is recommended for women at normal risk, beginning at age 40. In the U.K., women aged between 50 and 70 years are invited for breast screening every three years [4].

A mammogram screening is the most common and widely used technique for early detection of breast cancer. It is considered to be the most reliable and cost effective method for detection of breast cancer . In mammographic technique, a specialized low dose x-ray imaging modality is used to obtain a gray scale picture of breast region known as mammograms. Digital mammograms provide better dynamic contrast of breast tissues than screen film mammograms, and are widely utilized in CAD systems. A CAD system performs computerized mammographic analysis on digital mammograms to locate breast cancer. Now a days, many radiologist uses the results of CAD systems as a 2nd opinion before making a final decision. In general CAD techniques can be divided into two major stages as segmentation stage and computer aided cancer detection stage. In segmentation step, researchers perform segmentation on specific part of mammograms to extract breast region by removing noise, labels, markers and other artifacts. After complete extraction of breast region in mammogram, the next step involves removal of pectoral muscle from breast region. In second stage of CAD techniques, several texture features are extracted from normal and abnormal breast tissues and classifiers are trained via machine learning techniques in order to perform breast cancer detection in mammograms. Current study aims to introduce a simple methodology for segmentation of breast region, removal of pectoral muscle and detection of breast cancer in mammograms. Entire methodology is validated on Mini Mammographic Imaging Analysis Society (Mini-MIAS) database. The paper is organized as follows; section II explains previous techniques on automatic segmentation of breast region, removal of pectoral muscle and detection of breast cancer in mammograms. Section III discusses the proposed technique for automatic breast cancer detection. Section IV describe the results obtain with proposed technique and Section V compares the results of current methodology with previously developed methodologies. [2]

II. LITERATURE REVIEW

D.Dubey, S.Kharya, S.Soni worked on breast cancer prediction and stated that artificial neural networks are widely used. The paper featured about the advantages and short comings of using machine learning methods like SVM, Naive Bayes, Neural network and Decision trees [5]. Mustra et al. [6] proposed a robust and automatic technique for segmentation of breast region in which mammograms are first aligned and are then thresholded by threshold values obtained by k-means clustering in which total 10 clusters were formed. Afterwards, morphological operations were performed on binary mask to extract breast region. Mustra et al. [6] removed pectoral muscle by using standard edge detection technique and cubic polynomial estimation of muscle curvature. In this technique, 10 random points are selected from visible boundary of pectoral muscle which are then used for polynomial fitting of muscle boundary. Afterwards, cubic polynomial is used to estimate remaining invisible pectoral muscle boundary. ithya et al. [7] presented a performance comparison of three different feature extraction techniques for detection of breast cancer. A supervised neural network classifier was used to evaluate performance of Gray Level Co-occurrence Matrix (GLCM) features, intensity histogram features and intensity based features for detection of breast cancer in mammograms. Tai et al. [8] extracted GLCM and Optical Density Co-occurrence Matrix features. These features are classified by Linear Discriminant Analysis for detection of breast cancer. A. AlQoud et al. [9] used Gabor features and Local Binary Patterns features from normal and abnormal breast regions. A supervised Artificial Neural Network (ANN) based classification was performed to classify normal and abnormal breast tissues in mammograms. Vikas Chaurasia et al. [10] used three famous algorithms such as J48, Naive bayes, RBF, to build predictive models on breast cancer prediction and compared their accuracy. The results had shown that Naive Bayes predicted well among them with an accuracyof97.36%. Alireza Osarech, BitaShadgar [11] used SVM classification technique on two different benchmark datasets for breast cancer which got 98.80% and 96.63% accuracies. Haifeng Wang and Sang Won Yoon compared Naive Bayes Classifier, Support Vector Machine (SVM), AdaBoost tree, Artificial Neural Networks (ANN), to find a powerful model for breast cancer prediction. They implemented PCA for dimensionality reduction [12]. Nanayakkara et al. [13] proposed an automatic technique for breast region segmentation in mammograms. This technique employs a modified region growing technique known as fast marching technique for segmentation of breast muscle [1][2]. In proposed methodology, the mammogram images are pre-processed by using median filter and adaptive histogram equalization. From the mammogram images features are extracted using Gabor algorithm and also by calculating mean and standard deviation of the image. The selected features were then classified using Support Vector Machine classifier. SVM was trained to classify normal and abnormal breast tissues. The classifier first identifies whether the input image is normal or abnormal. If the image is identified to be abnormal means the breast masses are segmented from the pre processed images using Likelihood binarization algorithm. The segmentation algorithm segments the breast masses from the image based on the clustered result of group of pixels in the image. Then the mammogram image is classified into Benign or Malignant based on separate label and the features extracted.

III. METHODOLOGY

Block diagram as shown in Figure 1 represent the proposed methodology of our work. The mammogram images

were initially collected and the system is trained by extracting the features of the collected mammogram images based on Gabor algorithm and by calculating mean and standard deviation of the mammogram images. The Gabor algorithm captures a number of salient visual properties, including spatial localization, orientation selectivity, and spatial frequency selectivity. They are robust to illumination variations since they detect amplitude-invariant spatial frequencies of pixel gray values. The input mammogram image is taken and it is preprocessed using median filter. The median filter identifies the noisy pixels in the image and replaces it with the median value of the neighboring pixels. The preprocessed mammogram images are classified into normal or abnormal using SVM classifier based on the features extracted. The features were extracted from the image using Gabor algorithm and by calculating the mean and standard deviation of the image. If the image is identified to be abnormal image the image is segmented using Likelihood Binarization algorithm. The segmentation algorithm groups the similar pixels values in the image that are having similar intensities and pixel values. The center points of each cluster were calculated and the pixel value around each center value is specified in separate colors so that we are differentiating each clusters. Thus the breast masses were segmented based on the Likelihood binarization algorithm. The images are classified into benign or malignant based on the SVM classifier using different label and different test set. These steps are explained in more details in the following section.

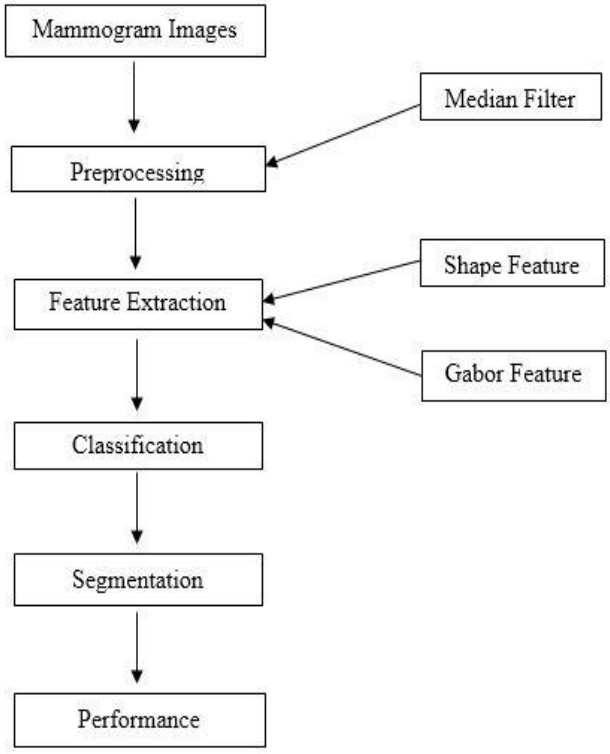


Figure. 1. Block diagram of proposed methodology

A. Median Filtering

Noise filtration is an important step in processing digital mammograms for CAD techniques. Most of mammograms usually contains low intensity noise near the skin-air interface of breast region. Sometimes, scratch artifacts are also present in mammograms. These noise and artifacts carries high frequency contents. Therefore, it is recommended to perform low pass filtering to suppress noise and artifacts in digital mammograms. Therefore, a non-linear median filter of window size 3× 3 was selected for noise suppression. A median filter run through every pixel of a mammogram, it takes a window of specified size from the neighborhood of each pixel say $I_{(x, y)}$ in image, sorts these pixels in ascending order and finally replaces the median value in sorted array with image pixel value $I_{(x, y)}$. The superiority of non-linear median filter over linear mean filter is that, it preserves the edges in an image and do not distribute noise content over its neighborhood pixels.

B. Feature Extraction

Once the noise is removed from breast muscle, some features were extracted from normal and abnormal breast tissues to express characteristics of cancerous tissues. Several feature extraction methods have been proposed as describe in literature review, but in our proposed methodology, features extraction was performed through GLCM. Total 14 features are extracted such as Mean, Standard Deviation and GLCM from preprocessed images. A GLCM describes occurrence of different combination of pixel intensities in an image.The Gabor feature, mean, Standard Deviation are extracted from the image. Gabor feature Capture a number of salient visual properties, including spatial localization, orientation selectivity, and spatial frequency selectivity, They are robust to illumination variations since they detect amplitude-invariant spatial frequencies of pixel gray values. The mean Standard Deviation values are calculated and the values are saved as features.

C. SVM Based Classification

In machine learning, support vector machines are supervised learning models with associated learning algorithms that analyze data and recognize patterns, used for classification and regression analysis. The basic SVM takes a set of input data and predicts, for each given input, which of two possible classes forms the output, making it a non-probabilistic binary linear classifier. Given a set of training examples, each marked as belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples into one category or the other. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same

space and predicted to belong to a category based on which side of the gap they fall on.

A SVM maps input vectors to a higher dimensional vector space where an optimal hyper plane is constructed. Among the many hyper planes available, there is only one hyper plane that maximizes the distance between itself and the nearest data vectors of each category. This hyper plane which maximizes the margin is called the optimal separating hyper plane and the margin is defined as the sum of distances of the hyper plane to the closest training vectors of each category.

Expression for hyper plane that does the separation is

$$W^T X + b = 0$$

x – Set of training vectors (input vector)

w – Vectors perpendicular to the separating hyper plane (adjustable weight vector)

b – Offset parameter which allows the increase of the margin (bias)

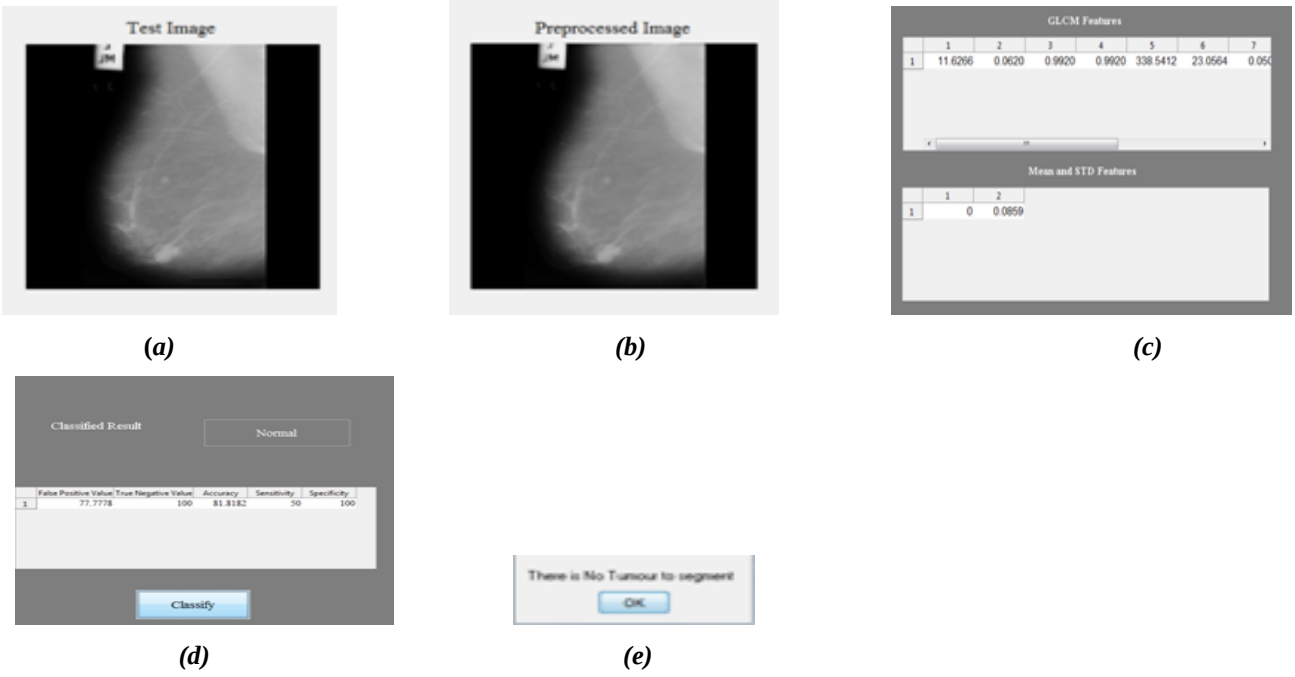
D. Segmentation

The ROI are selected from the image using Likelihood binarization algorithm. The segmentation is done based on the threshold specified. The regions within the threshold are grouped as a region and the regions that are different from the threshold are grouped into another region. The segmentation algorithm groups the similar pixels values in the image that are having similar intensities and pixel values. The center points of each cluster were calculated and the pixel value around each center value is specified in separate colors so that we are differentiating each clusters. Thus the breast masses were segmented based on the Likelihood binarization algorithm.

IV. RESULTS AND DISCUSSION

The development of the standard database files to analysis the breast cancer is the first stage of the process. Various standards dataset is used to develop the database for the detection of breast cancer. These were tested and have been validated with the known samples. These learning data and measurement is considered as a base and then actual samples are randomly authenticated. Samples including benign, malignant type breast cancer and normal images of breast has taken.

Case 1: Input image having no abnormalities.



(a) ***(b)*** ***(c)***

(d) ***(e)***

Figure. 2. (a) Test Image. (b) Preprocessed Image. (c) Feature Extraction Result. (d) SVM Classifier (e) No Tumor to segmentation message box

Case 2: Input image having tumor (Benign)

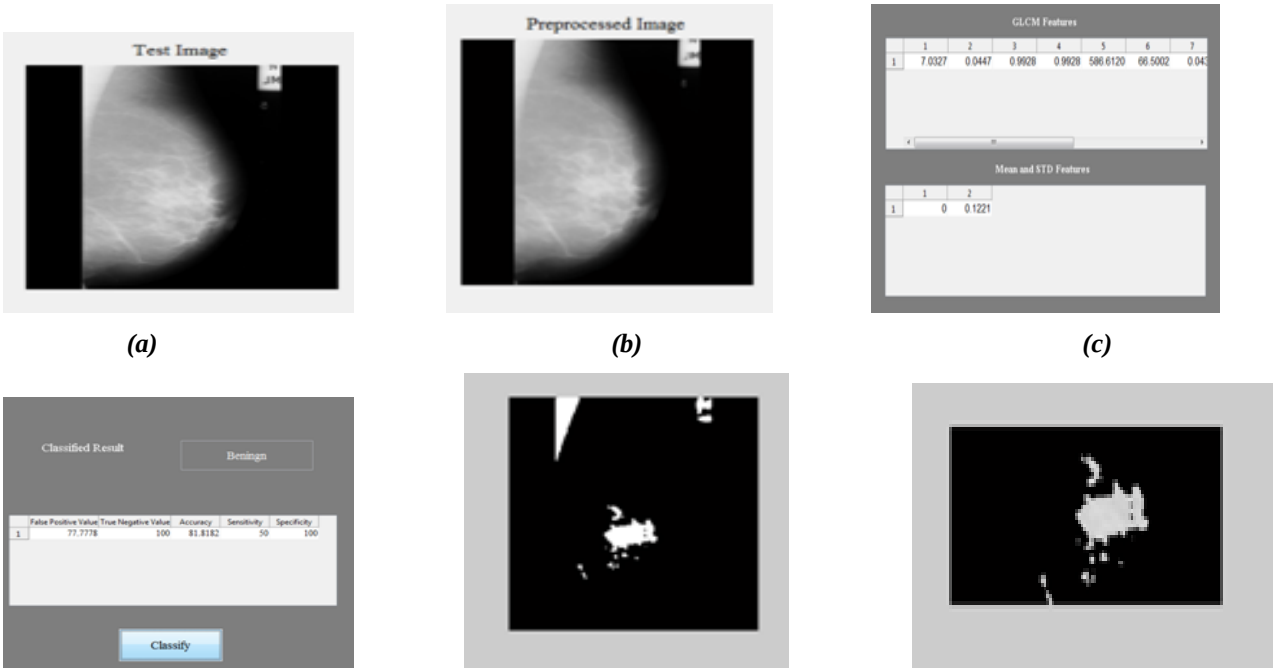


Figure. 3. (a) Test Image. (b) Preprocessed Image. (c) Feature Extraction Result. (d) SVM Classifier (e)Segmented pectoral muscle. (f) Removal of pectoral muscle and ROI

Case 3: Input Image having tumor (Malignant)

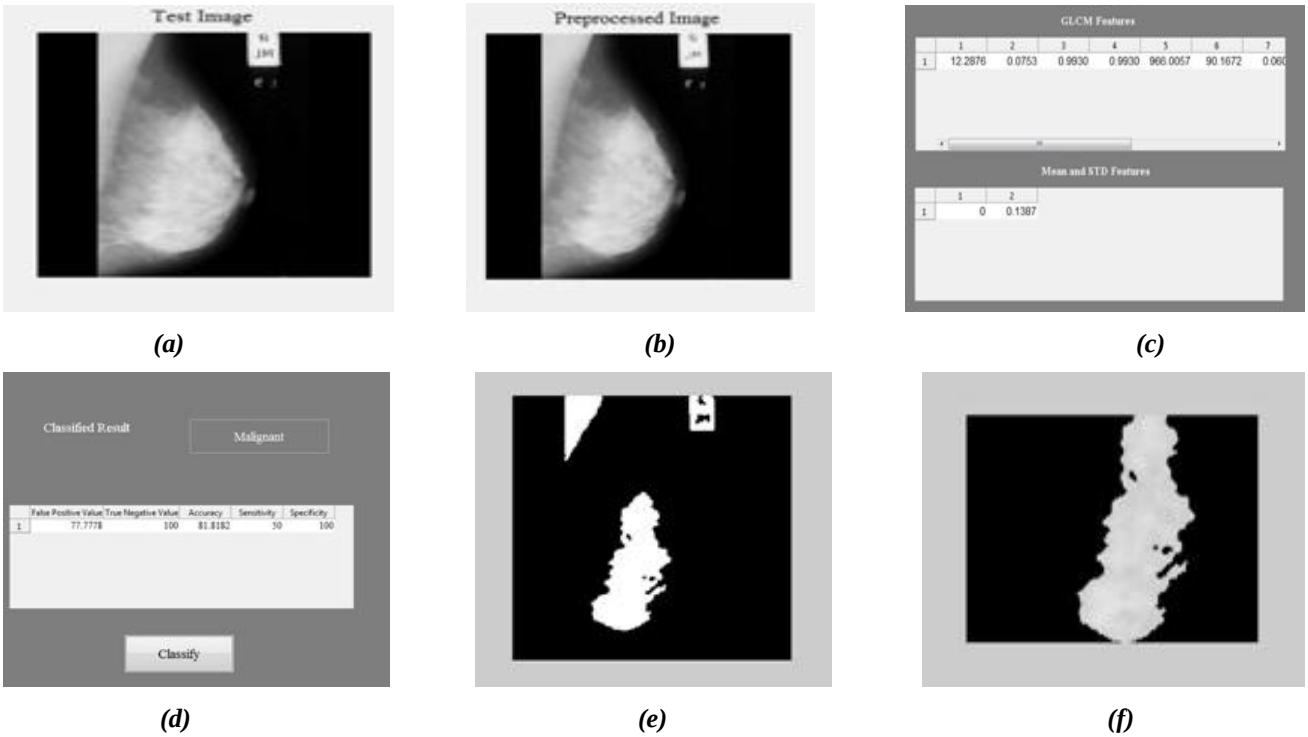


Figure. 4. (a) Test Image. (b) Preprocessed Image. (c) Feature Extraction Result. (d) SVM Classifier (e) Segmented pectoral muscle. (f) Removal of pectoral muscle and ROI

TABLE 1

COMPARISION BETWEEN PROPOSED WORK AND PREVIOUS WORKS

Sr.No	Method	Accuracy	Reference
1	Mean Shift segmentation	84%	Sultana et al., [14]
2	Discrete time Markov chain	84%	Wang et al., [15]
3	AD method	81%	Liu et al., [16]
4	Straight-line	85%	K.Vaidehi, et al., [17]
5	Geometry-based model	94%	Saeid Asgari Taghanaki et al., [18]
6	Homogenous texture and intensity deviation based method	92%	Li et al. [19]
7	Straight-line estimation and cliff detection	83.9%	Kwok et al. [20]
8	SVM Classifier	96%	Proposed

V. CONCLUSION

The proposed method has the potential to be used in CAD as a preprocessing step. Our experimental results also indicate that the proposed algorithm is versatile enough to be applied to extensive varieties in the appearance of the pectoral muscle. We present an algorithm that classifies the given image using the features extracted. In the preprocessing stage the noise in the images are removed. The ROI is chosen using the Likelihood binarization algorithm. The calculated Gabor features and the mean and the Standard deviation are the extracted features. Using the extracted features and the true label the SVM classifier classifies the image. The process can be further developed by employing some addition feature extraction algorithms such as LBP features and PCA features. The classifiers used can also be further developed by employing some other new classifiers that classifies the train features and produces better results.

REFERENCES

[1] Ch. Shravya, K. Pravalika, Shaik Subhani," Prediction of Breast Cancer Using Supervised Machine Learning Techniques,"International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-6, April 2019 1106 *Published By: Blue Eyes Intelligence Engineering & Sciences Publication Retrieval Number: F3384048619/19©BEIESP*

[2] Abdul Qayyum, A. Basit, "Automatic Breast Segmentation and Cancer Detection via SVM in Mammograms," 978-1-5090-3552-6/16/$31.00 _c 2016 IEEE

[3] IEEE TRANSACTIONS ON MEDICAL IMAGING, VOL. 34, NO. 2, FEBRUARY 2015 Analysis of Structural Similarity in Mammograms for Detection of Bilateral Asymmetry Paola Casti*, *Student Member, IEEE*, Arianna Mencattini, Marcello Salmeri, *Member, IEEE*, and Rangaraj M. Rangayyan, *Fellow, IEEE*

[4] IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING, VOL. 62, NO. 4, APRIL 2015 1203 Topological Modeling and Classification of Mammographic Microcalcification Clusters Zhili Chen∗, Harry Strange, Arnau Oliver, Erika R. E. Denton, Caroline Boggis, and Reyer Zwiggelaar

[5] D.Dubey ,S.Kharya, S.Soni and –"Predictive Machine Learning techniques for Breast Cancer Detection", International Journal of Computer Science and Information echnologies,Vol.4(6),2013,1023-1028.

[6] M. Mustra and M. Grgic, "Robust automatic breast and pectoral muscle segmentation from scanned mammograms," Signal processing, vol. 93, no. 10, pp. 2817–2827, 2013.

[7] R. Nithya and B. Santhi, "Comparative study on feature extraction method for breast cancer classification," Journal of Theoretical and Applied Information Technology, vol. 33, no. 2, pp. 1992–86, 2011.

[8] S.-C. Tai, Z.-S. Chen, and W.-T. Tsai, "An automatic mass detection system in mammograms based on complex texture features," IEEE journal of biomedical and health informatics, vol. 18, no. 2, pp. 618–627, 2014.

[9] A. AlQoud and M. A. Jaffar, "Hybrid gabor based local binary patterns texture features for classification of breast mammograms," International Journal of Computer Science and Network Security (IJCSNS), vol. 16, no. 4, p. 16, 2016.

[10] VikasChaurasia, BB Tiwari and Saurabh Pal – "Prediction of benign and malignant breast cancer using data miningstechniques",Journal of Algorithms and Computational Technology

[11] AlirezaOsarech, BitaShadgar,"A Computer Aided Diagnosis System for Breast Cancer",International Journal of Computer Science Issues, Vol. 8, Issue 2, March 2011

[12] Haifeng Wang and Sang Won Yoon – Breast Cancer Prediction using Data Mining Method, IEEE Conference paper

[13] R. Nanayakkara, Y. Yapa, P. Hevawithana, and P. Wijekoon, "Automatic breast boundary segmentation of mammograms," Int J. Soft Comput. Eng.(IJSCE), vol. 5, no. 1, 2015.

[14] Sultana, Alina, Mihai Ciuc, and Rodica Strungaru. "Detection of pectoral muscle in mammograms using a mean-shift segmentation approach." InCommunications (COMM), 2010 8th International Conference on, pp. 165-168. IEEE, 2010.

[15] Wang, Lei, Miao-liang Zhu, Li-ping Deng, and Xin Yuan. "Automatic pectoral muscle boundary detection in mammograms based on markov chain and active contour model," Journal of Zhejiang University SCIENCE C 11, no. 2 (2010): 111-118

[16] Liu, Li, Jian Wang, and Tianhui Wang. "Breast and pectoral muscle contours detection based on goodness of fit measure" In Bioinformatics and Biomedical Engineering,(iCBBE) 2011 5th International Conference on, pp. 1-4. IEEE, 2011.

[17] K.Vaidehi, T.S.Subashini, "Automatic Identification and Elimination of Pectoral Muscle in Digital Mammograms," *International Journal of Computer Applications (0975 – 8887) Volume 75– No.14, August 2013.*

[18] Saeid Asgari Taghanaki∗ , Yonghuai Liu , *Senior Member, IEEE*, Brandon Miles, and Ghassan Hamarneh, *Senior Member, IEEE, "*Geometry-Based Pectoral Muscle Segmentation From MLO Mammogram Views," IEEE transactions on biomedical engineering, vol. 64, no. 11, november 2017.

[19] Y. Li *et al.*, "Pectoral muscle segmentation in mammograms based on homogenous texture and intensity deviation," *Pattern Recognit.*, vol. 46, no. 3, pp. 681–691, 2013.

[20] S. M. Kwok *et al.*," Automatic pectoral muscle segmentation on mediolateral oblique view mammograms," *IEEE Trans. Med. Imag.*, vol. 23,no. 9, pp. 1129–1140, Sep. 2004.