IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727,
PP 01-06
www.iosrjournals.org

Detection and Classification of Breast Mass Using Support Vector Machine

Lothe Savita A.[1],Telgad Rupali L.[1],Siddiqui Almas.[1],Dr. Deshmukh Prapti D.[2]
[1]*Department Of Computer Science And It, Dr. B.A.M.U., Aurangabad, (Mh), India.*
[2]*Dr. G.Y. Pathrikar, College Of Cs & It, Aurangabad, (Mh), India.*

Abstract: *Accurately detecting the breast cancer disease in the early stage is extremely essential for fast recovery or to avoid the death probability. Breast cancer can be detected by various imaging modalities such as ultrasonography, magnetic resonance imaging, breast self-examination and mammography. Out of mammography is more used. Breast lesions are mass and microcalcification. In this paper we introduce the concept of mammography with its use to early detect the breast cancer, followed by CAD system including details about every step in it. CAD system can be used for classification of normal and abnormal tissue in the breast with the help of mammogram images. The main focus of this paper is to detect and classify the breast mass using support vector machine shows sensitivity 92.30%, specificity 62.50% with accuracy 86.84%.*
Keywords: *CAD, DFT, Enhancement,Mammogram, Segmentation, Seed region growing, SVM, Texture feature.*

I. Introduction

The term “breast cancer” refers to a malignant tumor that has developed from cells in the breast. Breast Cancer that forms in tissues of the breast, usually the ducts (tubes that carry milk to the nipple) and lobules (glands that make milk) [1].Breast cancer that forms in tissues of the breast, usually the ducts (tubes that carry milk to the nipple) and lobules (glands that make milk) [2]. Breast cancer is a leading cause of death among women in developed countries. The morbidity of breast cancer is increasing with a fast speed in developing countries due to the increase of life expectancy, urbanization and change in life styles [3]. According to Breast Cancer Statistics about 40,450 women in the U.S. are expected to die in 2016 from breast cancer [4]. As the cause of breast cancer is not clearly known, early detection remains the corner stone in breast cancer treatment. Breast cancer can be detected through various examinationas magnetic resonance imaging (MRI), mammography, ultrasound, CSE and BSE. Mammography is the most effective in reducing mortality rates by 30% - 70% [5].

Mammogram interpretation is a repetitive task which requires maximum care for avoidance of misinterpretation. Therefore, computer aided detection and diagnosis (CAD) system is currently a very popular and efficient method which analyses the digital mammogram with the use of image processing. CAD system helps radiologists in accurate interpretation of mammograms for detection of suspicious lesions and classification. It has been observed that 60%-90% of the biopsies of cancers predicted by radiologists found benign-malignant mammogram. Computer aided detection system is a combination of image-processing techniques and intelligent methods that can be used to enhance the medical interpretation process can make better results in the development of more efficient diagnosis. The computer outcome assists radiologists in image analysis and diagnostic decision making. In addition, a CAD system could direct a radiologist’s attention to regions where the probability of indications of disease is greatest. A CAD system provides reproducible and quite realistic outcomes [6].

This work is organized as follows. In section II provides literature review of detection and classification techniques. Section III gives about CAD system, followed by database used for experimental purpose in IV section. Section V gives the details of proposed system and its result.

II. Background

This section gives the literature review about the various detection and classification techniques.
Faye et al [7], proposes method use for classification of images is based on preselecting features based on their capabilities of differentiating classes using a T test. Random subsets achieving a predefined accuracy rate are then used to generate a final set of features. The method was used in this work with wavelet transform with LDA and kNN classifiers. Although the final accuracy rate obtained in the experiments are relatively low, the improvement when combining classifiers is highly encouraging.

Pereira D. C. et al [8] presents a set of computational tools to aid segmentation and detection of mammograms that contained mass or masses in CC and MLO views. An artifact removal algorithm is first implemented followed by an image denoising and gray-level enhancement method based on wavelet transform and Wiener filter. Finally, a method for detection and segmentation of masses using multiple thresholding, wavelet transform and genetic algorithm is employed in mammograms which were randomly selected from the Digital Database for Screening Mammography (DDSM).

Jen C. et al [9], proposed a high-performance CAD system for detecting abnormal mammograms by using the two-stage classifier ADC, which applied the PCA-based technique accompanied by robust feature weight adjustments.

III. Cad System

CAD is invented to follow the subsequent steps:
Initially mammogram images read by radiologists for marking suspicious areas.
A CAD system scanning to detect suspicious features.
Radiologists analysis of the prompts given by the CAD system and verification if the suspicious areas were left unchecked in the first reading [10].

1.1 Steps in Computer Aided Detection

For the time being, mammographic screening remains the most effective method for early detection of breast cancer. However, reading mammography is a time-consume error prone work. Therefore, many computer-aided detection and diagnosis systems (CAD) have been developed to assist radiologists in detecting and classifying mammographic lesions [11]. Figure 1 gives the steps in CAD system.

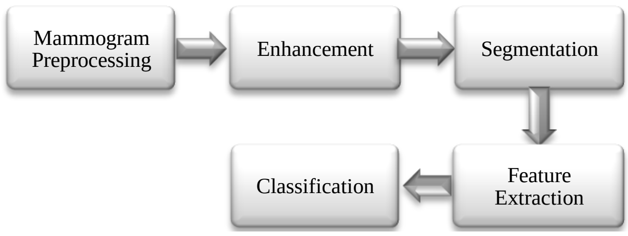


Fig 1: Basic steps in Computer Aided Detection (CAD) System

IV. Database

All mammograms used in this work (detection of mass) are from a mini Mammographic database provided by Mammographic Image Analysis Society (MIAS), includes radiologist’s “truth” markings on the locations of any abnormalities that may be present [12-13].

It contains total 322 images (Medio-Lateral Oblique (MLO)) representing 161 bilateral pairs. The database is divided into seven categories. These include normal image pairs and abnormal pairs containing microcalcifications, circumscribed masses, spiculated lesions, ill-defined masses, architectural distortion and asymmetric densities. Each mammogram from the database is a 1024 X 1024 pixels and with a spatial resolution of 200 m/pixel.

V. Proposed System

Mass is one of the main signs of breast cancer. A mass is a space occupying lesion which usually exhibits higher intensity compared with fat tissues in a mammogram. The diction between normal mass has been largely difficult task for the radiologists, as masses can be subtle at a nearly stage, shapes and sizes [3].
A mass is defined as a space-occupying lesion seen in more than one projection. A mass with a regular shape has a higher probability of existence benign, whereas a mass with an irregular shape has a higher probability of presence malignant [14-15].
Figure 2 shows the flow of proposed segmentation of mass using seeded region growing method.

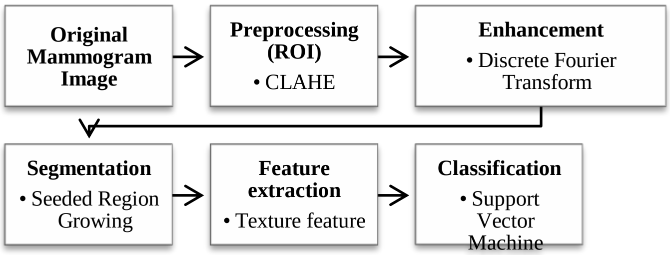


Fig 2: Proposed sytem for segmentation of Mass

Here we developed CAD system for detection of mass consisting image processing techniques as preprocessing, enhancement, segmentation, feature extraction and classification.

Mammogram image is composed of different noises, artifacts in their background. The object area also contains the pectoral muscles. All these areas are unwanted portions for the texture analysis due to which the full mammographic image is unsuitable for feature extraction and subsequent classification. Therefore, a cropping operation has been applied on mammogram images to extract the regions of interests (ROIs) which contain the abnormalities, excluding the unwanted portions of the image. This process is performed by referring the center of the abnormal area as the center of ROI and taking a square enclosing the abnormal area [16] shown in figure 3.3. ROI is a fundamental step used as input for the developed CAD system, representing the selected region of the mammographic image.

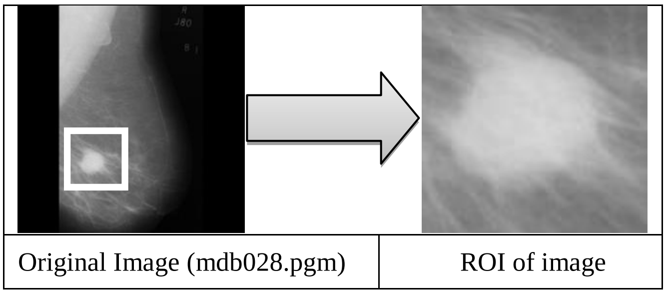


Fig 3: ROI selecting from image

5.1 Preprocessing

Mammograms are diffcult images to interpret and a preprocessing phase is necessary to improve the quality of the images [17]. Here the preprocessing stage constists two main stages. In this system the first phase involves the removal of background information and noise from the image with the help of median filters, while the second phase deals with enhancing the contrast of interest areas using CLAHE (Contrast Limited Adaptive Histogram Equalization) technique. CLAHE algorithm operates on small regions of the gray image by partitioning the image into contextual regions called tiles [15].

5.2 Image Enhancement

Image enhancement is the process of adjusting digital images so that the results are more suitable for display or further analysis. We remove noise or brighten an image, making it easier to identify key features [18]. Here we used Discrete Fourier Transform (DFT) technique for enhancement of image shown in figure 4.

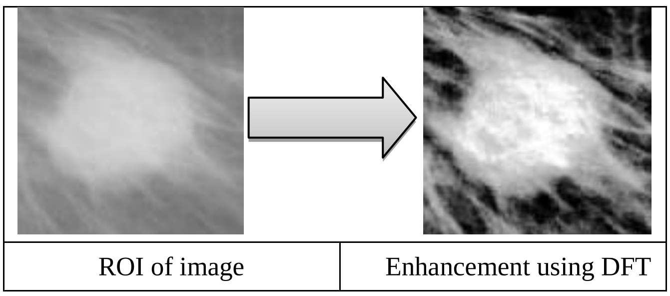


Fig 4: Enhancement using DFT

Fourier transform is a continuous signal can be represented as a (countable) weighted sum of sinusoids. It offers flexibility in the design and implementation of filtering solutions in areas such as image enhancement. The Fourier transform has a number of interesting properties [19].

Following are the steps in DFT filtering

1. Convert the image to floating point.
2. Obtain the padding parameters.
3. Obtain the Fourier transform.
4. Generate the highpass filter function.
5. Multiply the transform by the filter.
6. Obtain the inverse FFT.
7. Crop the top, left rectangle to the original size.
8. Convert the filtered image to the class of input image [20-21].

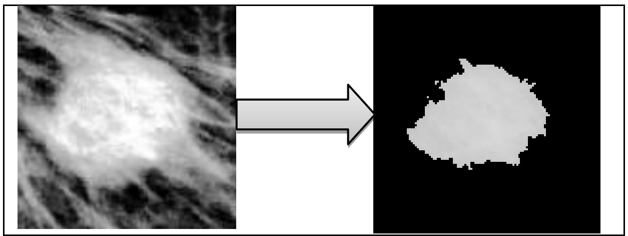
5.3 Segmentation

Segmentation methods broadly divided into two main categories: edge-based methods and region-based. In edge-based methods, the local discontinuities are detected first and then connected to form longer, hopefully complete, boundaries. In region-based methods, areas of an image with homogeneous properties are found, which in turn give the boundaries [22].

In this work we have used region growing method for the segmentation of mass. Region-growing techniques exploit the important fact that pixels which are close together have similar gray values. The algorithm starts from one or more pixels, called ***Seeds,*** chosen as the reference values for the growth. Then, region growing expands the selected areas around the seeds to include nearby pixels falling within a threshold range. The main steps are summarized as follows.

1. Choose the seed pixels.
2. Check the neighboring pixels and add them to the region if they are similar to the seed.
3. Repeat step 2 for each of the newly added pixels; stop if no more pixels can be added [23].

The result of segmentation is shown in figure 6.



Enhanced Image	Segmented region

Fig 4: Segmented Region

5.4 Feature Extraction

Analysis of biomedical images requires the extraction of numerical features that characterize most significant properties of region of interest [24].

Texture is a pattern of the occurrence of gray levels in image space, to measure the random texture, in this work we have used 5 shape features and 14 Haralick's statistical measures proposed by Haralick et al [25] of the segmented region given in fig 4. Haralick's measures are based upon the moments of a joint PDF that is estimated as the joint occurrence or co-occurrence of gray levels, known as the gray-level co-occurrence matrix (GCM), GCMs are also known as spatial gray-level dependence (SGLD) matrices [24].

5.5 Classification

Once the features of segmented image related to masses are extracted and selected, the features are given as input to classifier for classification benign masses, or malignant masses. The SVM is based on the idea of hyper-plane classifier, and it tries to look for the hyper-plane that maximizes the margin between two classes [26-27]. In this work, we use SVM to classify masses as benign or malignant.

SVM is a supervised learning methods used for classification and regression analysis. SVM is a binary linear classifier that for each given input data it predicts which of two possible classes comprises the input. It was found that for identification of mass SVM shows the accuracy 86.84%, with sensitivity 92.30% and specificity 62.50% on mini-mias database.

VI. Experimental Result

After segmentation stage the feature that are extracted from image. Extracted features are further given to SVM classifier for classification of masses into benign or malignant. The performance analysis of system is done with accuracy, sensitivity, specificity.

Accuracy = (TP + TN) / Total number of samples-------- (1)

True Positive Rate (Sensitivity) = TP / (TP + FN)----------(2)

True Negative Rate (Specificity) = TN / (TN + FP)--------(3)

where

- True Positive (TP), when the suspected abnormality is in fact malignant;
- True negative (TN), when there is no detection of abnormality in a healthy person;
- False positive (FP), when occurs detection of abnormality in a healthy person;
- False negative (FN), when there is no detection of a malignant lesion.

VII. Conclusion

Mammography screening remains the most effective method for early detection of breast cancer. However, reading mammography is a time-consume error work. Therefore, many computer-aided detection and diagnosis systems (CAD) have been developed to assist radiologists in detecting and classifying mammographic lesions. A mass screened on a mammogram can be either benign or malignant depending on its shape. Benign tumors usually have round or oval shapes, although malignant tumors have a partially rounded shape with a spiked or irregular outline. Non-cancerous or benign tumors include cysts, fibro adenomas, and breast hematomas. A cancerous or malignant tumor in the breast is a mass of breast tissue that grows in an abnormal and uncontrolled way. Normally, malignant masses appear brighter than any tissue surrounding it. In this work we proposed CAD system for detection and classification of breast mass. Further extracting the statistical values of segmented region classification is performed to find breast tumor is benign or malignant.

References

[1]. Prof. Samir Kumar Bandyopadhyay, IndraKanta Maitra, Souvik Banerjee, "Digital Imaging in Pathology Towards Detection and Analysis of Human Breast Cancer", IEEE Second International Conference on Computational Intelligence, Communication Systems and Networks, July-2010, pp. 295-300, doi: 10.1109/CICSyN.2010.43.

[2]. American Cancer Society, "What is Cancer", [Online], Available: http://www.cancer.org/Cancer/CancerBasics/what-is-cancer, American Cancer Society, ACS.

[3]. Yanfeng Li, Hougin Chen, Yongyi Yang, Lin Cheng, Lin Cao, "A bilateral analysis scheme for false positive reduction in mammogram mass detection", Computers in Biology and Medicine, vol. 57, pp. 84-95, Elsevier, 2015.

[4]. U.S. Breast Cancer Statistics, Breastcancer.org, http://www.breastcancer.org/symptoms/understand_bc/statistics

[5]. K. P. Lochanambal, M. Karnan, R. Sivakumar, "Identifying Masses in Mammograms Using Template Matching", Second International Conference on Communication Software and Networks, IEEE, 2010.

[6]. FarzanKhatib, Rozi Mahmud, SyamsiahMashohor, M. Iqbal Saripan, Raja SyamsulAzmir Raja Abdullah, "Automated Cystic Mass Extraction from Ultrasound Phantom Images", Sixth Asia Modelling Symposium, pp. 54-58, IEEE, 2012.

[7]. Ibrahima Faye, "A random feature selection method for classification of Mammogram images", Third International Conference on Intelligent Systems Modelling and Simulation, ISMS, pp 330-333, IEEE, 2012.

[8]. Danilo Cesar Pereira, Rodrigo Pereira Ramos, Marcelo Zanchetta do Nascimento, "Segmentation and detection of breast cancer in mammograms combining wavelet analysis and genetic algorithm", Computer Methods and Programs in Biomedicine 114, Elsevier, 2014, pp 88-101.

[9]. Chun-Chu Jen, Shyr-Shen Yu, "Automatic detection of abnormal mammograms in mammographic images", Expert Systems with Applications 42, Elsevier, 2015, pp 3048-3055.

[10]. Rangayyan, R., Ayres, F., &Desautels, J., "A review of computer-aided diagnosis of breast cancer: Toward the detection of subtle signs", Journal of the Franklin Institute, vol. 344, pp. 312-348, 2007.

[11]. Yihua Lan, Haozheng Ren, Jinxin Wan, "A Hybrid Classifier for Mammography CAD", Fourth International Conference on Computational and Information Sciences, pp. 309-312, IEEE, 2012.

[12]. J. Suckling, "The Mammographic Image Analysis Society Digital Mammogram Database", ExerptaMedica, International Congress Series 1069, pp. 375-378, 1994.

[13]. Mohammed J. Islam, Majid Ahmadi, and Maher A. Sid-Ahmed, "Computer Aided Detection and Classification of Masses in Digitized Mammograms Using Artificial Neural Network", ICSI 2010, Part II, LNCS 6146, Springer, pp. 327–334, 2010.

[14]. Jinshan Tang, Rangaraj M. Rangayyan, Jun Xu, Issam El Naqa, and Yongyi Yang,"Computer aided detection and diagnosis of breast cancer with mammography: Recent advances", IEEE Transactions on Information Technology in

[15]. Nijad Al-Najdawi, Mariam Biltawi, Sara Tedmori, "Mammogram image visual enhancement, mass segmentation and classification", Applied Soft Computing 35, pp. 175-185, Elsevier, 2015.

[16]. ShradhanandaBeura, BanshidharMajhi, Ratnakar Dash, "Mammogram classification using two dimensional discrete wavelet transform and gray-level co-occurrence matrix for detection of breast cancer", Neurocomputing 154, Elsevier, 2015, pp. 1-14.

[17]. Maria Victoria Carreras Cruz, Patricia Rayon Vilella, "Circumscribed Mass Detection in Digital Mammograms", Proceedings of the Electronics, Robotics and Automotive Mechanics Conference (CERMA), 2006
[18]. Prakash Bethapudi, Dr. E. Srinivasa Reddy, Dr.Madhuri. P.,"Detection of Malignancy in Digital Mammograms from Segmented Breast Region Using Morphological Techniques", IOSR Journal of Electrical and Electronics Engineering (IOSR-JEEE), Volume 5, Issue 4, pp 9-12, June 2013.
[19]. http://www.fpferrarese.com/images0910/MISC/Digital_Imag_Processing_Medical_Applications_Frequency_Domain.pdf
[20]. Rafael C. Gonzalez, Richard E. Woods, "Digital Image Processing", Third Edition, Pearson., Rafael C. Gonzalez, Richard E. Woods, Steven L. Eddins, "Digital Image Processing using Matlab", Second Edition, McGraw Hill.
[21]. Lothe Savita A., GaikeVrushali V., Dr.DeshmukhPrapti D., "Visual Enhancement of mammogram image using DFT", Indian Science Congress – 2015, University of Mumbai, 3 – 7 Jan 2015.
[22]. Subarna Chatterjee, Ajoy Kumar Ray, Rezaul Karim, Arindam Biswas, "Detection of micro-calcification to Characterize Malignant Breast Lesion", Third National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics, pp. 251-254, IEEE, 2011.
[23]. Arianna Mencattini, Giulia Rabottino, Marcello Salmeri, Roberto Lojaconoand Emanuele Colini, "Breast Mass Segmentation in Mammographic Images by an Effective Region Growing Algorithm", ACIVS 2008, LNCS 5259, pp 948–957, Springer, 2008.
[24]. Rangaraj M. Rangayyan, "Biomedical Image Analysis", The Biomedical Engineering Series, CRC Press, 2005.
[25]. Robert M. Haralick, K. Shanmugam and Its'HakDinstein, "Textural Features for Image Classification", IEEE transactions on Systems, Man and Cybernetics, Vol. No-6, Nov. 1973.
[26]. Chuin-Mu Wang, Xiao-Ding Mai, Geng-Cheng Lin, Chio-Tan Kuo, "Classification for Breast MRI Using Support Vector Machine", International Conferences on Computer and Information Technology Workshops, IEEE, pp 362-367, 2008
[27]. Savita A. Lothe, Shobha K. Bawiskar, Rupali P. Moharkar, Dr.Prapti D. Deshmukh, "A Survey of Image Processing Algorithms for Detecting Microcalcification in Mammogram Images", International Journal of Advanced Research in Computer Science, Volume 4, No. 1, pp 1-3, January 2013.