

DETECTION OF BREAST CANCER USING BPN CLASSIFIER IN MAMMOGRAMS

Brundha.k^[1],Gali Snehapriya^[2],Swathi.U^[3],Venkata Lakshmi.S^[4]

Abstract: This paper describes a computer-aided detection and diagnosis system for breast cancer, the most common form of cancer among women, using mammography. The system relies on the Multiple-Instance Learning (MIL) paradigm, which has proven useful for medical decision support in previous works from our team. In the proposed framework, breasts are first partitioned adaptively into regions. The GLCM Features are extracted from wavelet sub bands. Then, features derived from the detection of lesions (masses and micro calcifications) as well as textural features, are extracted from each region and combined in order to classify mammography examinations as “normal” or “abnormal”. Whenever an abnormal examination record is detected, the regions that induced that automated diagnosis can be highlighted. Two strategies are evaluated to define this anomaly detector. In a first scenario, manual segmentations of lesions are used to train an NN that assigns an anomaly index to each region; local anomaly indices are then combined into a global anomaly index.

keywords: Computer-aided diagnosis, Grey level co-occurrence Matrix (GLCM), Back Propagation Network (BPN), Wavelet Decomposition.

1. INTRODUCTION

Breast cancer is the second most common cancer in the world and more prevalent in the female population. This is the second leading cause of death for women all over the world. At the international level, it represents nearly 22%

of new cancer cases and the 5-year survival is 61% globally. According to the WHO (World Health Organization) breast cancer causes 450,000 deaths worldwide each year [1].

Mammography remains the most effective and valuable tool of detection of breast abnormalities and many applications in the literature proved its effective use in breast cancer diagnosis. X-ray mammography is currently known as the most cost-effective imaging modality for the early detection of breast cancer, and thus, mammograms are obtained regularly in the breast screening program.

A huge number of mammograms are taken by the breast screening program and these mammograms are visually examined by experts to detect the signs of abnormalities. The sensitivity of mammograms varies between approximately 70% and 90%, depending on the following factors: size and location of the lesion, density of the breast tissue, patient age, exam quality and the radiologists interpretation ability [1][2][3].

Breast calcifications are deposits of calcium within the soft tissue [2][7]. There are two types: macro calcifications and micro calcifications. Micro calcifications (MCs) are tiny deposits of calcium salts which can be located anywhere in breast tissue. Although mammography is considered the most effective screening tool for the examination of breast MCs [2], specific

inherent limitations of the method led to the development of Computer Aided Diagnosis (CADx) systems.

The role of a CADx system is to help the radiologists in their diagnostic process by supporting them with a reliable second opinion. Each CADx system includes specific steps such as the segmentation of micro calcifications, the feature extraction process and the final classification of a considered case [4].

Computer-aided diagnosis and detection (both usually referred to as CAD) systems report a high level of sensitivity in revealing mostly lesions and calcifications but they typically fall short of detecting architectural distortion (AD) of breast tissue—the third most common sign of impalpable breast cancer.

The clustering of micro calcification indicates a cancer. Thus, micro calcification Clusters (MCCs) on a mammogram is one of the most important signs of breast cancer. However, it is not easy to detect micro calcification due to their small size, low contrast, and similarity to the dense tissue in the mammogram. A new scheme for the computer-aided diagnosis (CAD) of micro calcification clusters (MCCs) detection in a MultiInstance Learning (MIL) framework is proposed in this paper. In the Multi-Instance Learning framework, each training example is regarded as a bag of instances. A bag is positive if it contains at least one positive instance, and otherwise negative. The method proposed in this paper to detect malignant tumors consists of a two-step process. The first step is intended to detect tumor candidates. The second step is an evaluation of their malignancy in order to reduce the number of false positives.

3. RELATED WORKS

paper presents the approaches which are applied to develop CAD systems on mammography and ultrasound

images. The performance evaluation metrics of CAD systems are also reviewed.[1].Algorithm concurrently delineates the boundaries of the breast boundary, the pectoral muscle, as well as dense regions that include candidate masses[2].

The Digital Database for Screening Mammography (DDSM) was used in this work for the acquisition of mammograms[3]. The sensitivity remains at high levels, while improving both the accuracy, from 51.4% to 69%, and the specificity, from 16.6% to 54.7%[4].

The proposed two-layer architecture first collects low-level rotation-invariant textural features at different scales and then learns latent textural primitives from the collected features by GMM[5]. The same algorithm is used whatever the dimensionality of the signal, and whatever the lattice used[6]. Features are extracted from the potential candidates based on a constructed graph[7].

It introduces spot enhancement using CWT properties. Automatically changes image resolution by converting scales from millimeter to pixel resolution[8]. CR images is used to test the performance. 96% of malignant tumors are detected[9]. The optimized wavelets achieved a sensitivity of 90% with a specificity of 80%, whereas the conventional wavelets achieved a sensitivity of 80% with the same specificity [10].

4. METHODOLOGY USED

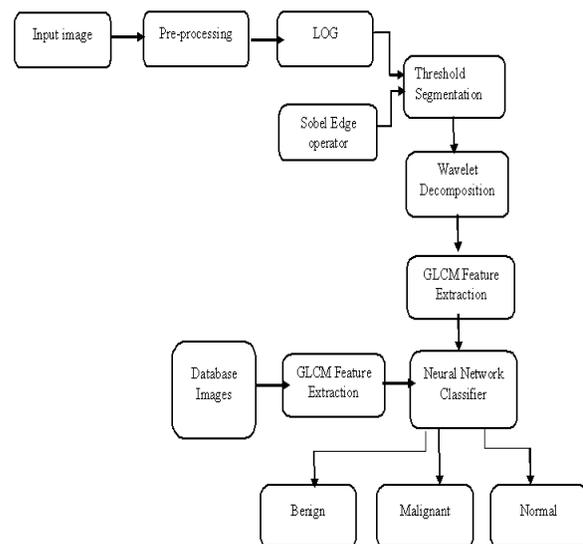


FIG.4.1. ARCHITECTURE DIAGRAM

a. preprocessing: Image pre-processing is the term for operations on images at the lowest level of abstraction. These operations do not increase image information content but they decrease it if entropy is an information measure. The aim of pre-processing is an improvement of the image data that suppresses undesired distortions or enhances some image features relevant for further processing and analysis task.

b. GLCM :

The GLCM Features are extracted from wavelet sub bands. The, features derived from the detection of lesions (masses and micro calcifications) as well as textural features, are extracted from each region and combined in order to classify mammography examinations as “normal” or “abnormal”.

GLCM properties:

i) Co-occurrence Matrix: A Co-occurrence matrix (CCM) by calculating how often a pixel with the intensity (gray-level) value *i* occurs in a specific spatial relationship to a pixel with the value *j*. By default, the spatial relationship is defined as the pixel of interest and the pixel to its immediate right (horizontally adjacent), but you can specify other spatial relationships between the two pixels. Each element (*i,j*) in the resultant ccm is simply the sum of the number of times that the pixel with value *i* occurred in the specified spatial relationship to a pixel with value *j* in the input image. The number of gray levels in the image determines the size of the CCM.

ii) Energy: It is a measure the homogeneousness of the image and can be calculated from the normalized COM.

$$J = \sum_{i=1}^n \sum_{j=1}^n (p(i, j))^2$$

iii) Entropy : Entropy gives a measure of complexity of the image. Complex textures tend to have higher entropy .

$$S = - \sum_{i=1}^n \sum_{j=1}^n p(i, j) \log(p(i, j))$$

Where, *p(i, j)* is the co occurrence matrix

iv) Contrast : Measures the local variations and texture of shadow depth in the gray level co-occurrence matrix.

$$\text{Contrast} \quad I = \sum \sum (x-y)^2 p(x, y)$$

v) Correlation : Measures the joint probability occurrence of the specified pixel pairs.

$$\text{Cor:} \sum(\sum((x-\mu_x)(y-\mu_y)p(x, y)/\sigma_x\sigma_y))$$

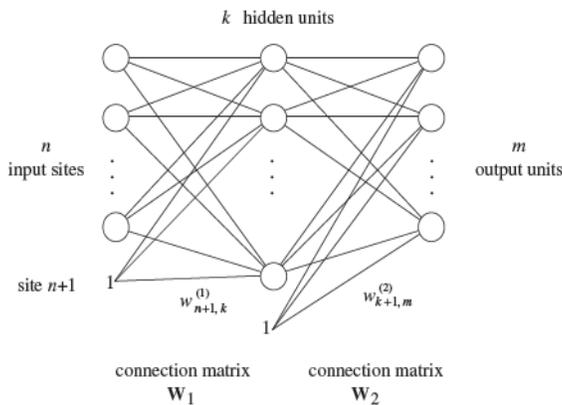
vi) Homogeneity : Measures the closeness of the distribution of elements in the GLCM to the GLCM diagonal.

$$\text{Homo :} \sum(\sum(p(x, y)/(1 + |x-y|)))$$

c. Back Propagation Network:

The performance of the Back Propagation network was evaluated in terms of training performance and classification accuracies. Back Propagation network gives fast and accurate classification and is a promising tool for classification of the tumors. Back propagation algorithm is finally used for classifying the pattern of malignant and benign tumor. The back-propagation learning rule can be used to adjust the weights and biases of networks to minimize the sum squared error of the network. It is used to compute the necessary corrections, after choosing the weights of the network randomly.

FIG.4.2.NEURAL NETWORK



The algorithm can be decomposed in the following four steps:

- i) Feed-forward computation
- ii) Back propagation to the output layer
- iii) Back propagation to the hidden layer
- iv) Weight updates

The algorithm is stopped when the value of the error function has become Sufficiently small. The following figure is the notation for three layered network.

d. Classification:

The BPN with FF is trained with reference features set and desired output using 'newff' and 'train' command. Here, target 1 for dataset1, 2 for dataset2 and dataset3 are taken as desired output. After the training, updated weighting factor and biases with other network parameters are stored to simulate with input features. At the classification stage, test image features are utilized to simulate with trained network model using 'sim' command. Finally it returns the classified value as 1 ,2 or 3 based on that the decision will be taken as Normal, Benign or Malignant.

e.Performance Metrics:

The performance of classifier can be evaluated through following parameters,

i)Sensitivity: It measures the proportion of actual positives which are correctly identified.

$$\text{Sensitivity} = \frac{Tp}{(Tp + Fn)}$$

Where, $Tp = \text{True Positive}$: Abnormality correctly classified as Abnormal.

$Fn = \text{False negative}$: Abnormality incorrectly classified as normal

ii)Specificity:

It measures the proportion of negatives which are correctly identified.

$$\text{Specificity} = \frac{Tn}{(Fp + Tn)}$$

Where, $Fp = \text{False Positive}$: Normal incorrectly classified as Abnormal

$Tn = \text{True negative}$: Normal correctly classified as normal.

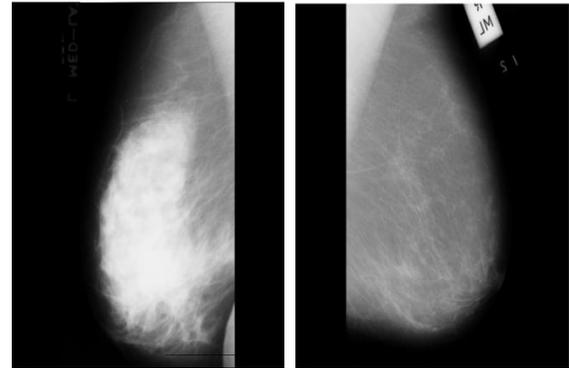
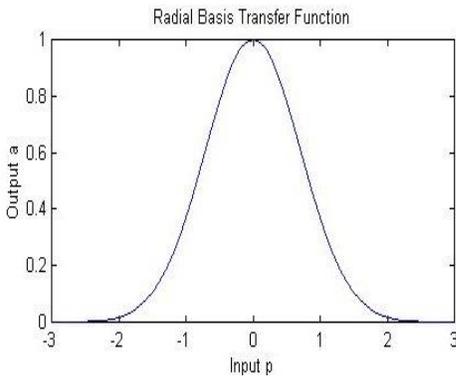
$$\text{Totalaccuracy} = \frac{(Tp+Tn)}{(Tp+Tn+Fp+Fn)}$$

The network generates the performance metrics, *Sensitivity: 100%, Specificity: 75%, Accuracy: 90.9091%.*

5. EXPERIMENTAL RESULTS

Radial Basis Function :

Different types of radial basis functions could be used, but the most common is the Gaussian function:



C. Malignant stage d. Normal stage

Normal versus Abnormal Examination Records: for the purpose of this study, examination records were divided into two groups: 'abnormal' examination records, which correspond to examination records with at least one benign or one cancerous finding, and 'normal' examination records, which correspond to examination records with no findings. Statistics are reported in table I. Statistics about the DDSM dataset

Diagnosis	DBA	HOWTEK	LUMISYS
normal	430	183	82
abnormal	97	966	721
Total	527	1149	803

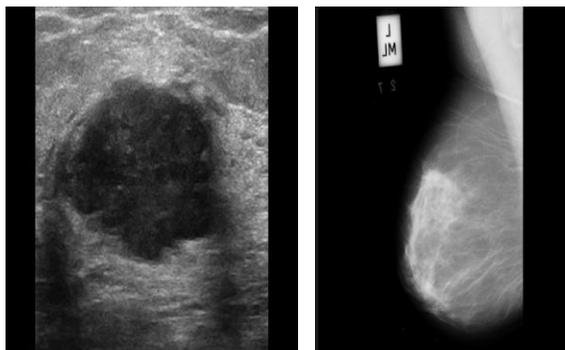
6.CONCLUSION

In this paper, we detect the cancer stages as benign, malignant and normal. We use LOG and Sobel Edge Operator to extract the shape and edge features. Noise is avoided using threshold segmentation. In this wavelet transform is used to recognize the cancer tissues. The GLCM feature is used for representing highly affected area. Using neural network classifier, we will compare the breast images in the database with the input image and identify the cancer stages as benign, malignant and normal. This method identifies accurately the stages of breast cancer. It can segment the cancer regions from the image accurately. It is useful to classify the cancer images for accurate detection. Early stage Detection of cancer from images is possible using BPN classifier.

7. REFERENCES

[1] K. Ganesan, U. R. Acharya, C. K. Chua, L. C. Min, K. T. Abraham, and K. H. Ng, "Computer-aided breast cancer detection using mammograms: a review," IEEE Rev Biomed Eng, vol. 6, pp. 77– 98, 2013.

[2] A. Jalalian, S. B. Mashohor, H. R. Mahmud, M. I. Saripan, A. R. Ramli, and B. Karasfi, "Computer-aided detection/diagnosis of breast cancer in



a. cancer affected breast b. Benign stage

mammography and ultrasound: a review," *Clin Imaging*, vol. 37, no. 3, pp. 420–6, May–Jun 2013.

[3] J. Melendez, B. van Ginneken, P. Maduskar, R. H. H. M. Philipsen, K. Reither, M. Breuninger, I. M. O. Adetifa, R. Maane, H. Ayles, and C. I. S´anchez, "A novel multiple-instance learning-based approach to computer-aided detection of tuberculosis on chest X-rays," *IEEE Trans Med Imaging*, vol. 34, no. 1, pp. 179–92, Jan. 2015.

[4] C. Li, K. M. Lam, L. Zhang, C. Hui, and S. Zhang, "Mammogram microcalcification cluster detection by locating key instances in a multi-instance learning framework," in *Proc IEEE ICSPCC*, 2012, pp. 175–9.

[5] C. R. Maurer Jr., R. Qi, and V. Raghavan, "A linear time algorithm for computing exact Euclidean distance transforms of binary images in arbitrary dimensions," *IEEE Trans Pattern Anal Mach Intell*, vol. 25, no. 2, pp. 265–70, Feb 2003.

[6] B. W. Hong and B. S. Sohn, "Segmentation of regions of interest in mammograms in a topographic approach," *IEEE Trans Inf Technol Biomed*, vol. 14, no. 1, pp. 129–39, Jan 2010.

[7] G. Loy and A. Zelinsky, "Fast radial symmetry for detecting points of interest," *IEEE Trans Pattern Anal Mach Intell*, vol. 25, no. 8, pp. 959–73, Aug 2003.

[8] R. M. Haralick, K. Shanmugam, and I. D. Dinstein, "Textural features for image classification," *IEEE Trans Syst Man Cybern*, vol. SMC-3, no. 6, pp. 610–21, Nov 1973.

[9] R. Venkatesan, P. Chandakkar, B. Li, and H. K. Li, "Classification of diabetic retinopathy images using multi-class multiple-instance learning based on color correlogram features," in *Proc IEEE EMBS*, vol. 2012, 2012, pp. 1462–5.

[10] "Adaptive nonseparable wavelet transform via lifting and its application to content-based image retrieval," *IEEE Trans Image Process*, vol. 19, no. 1, pp. 25–35, Jan. 2010.

[11] G. Quellec, M. Lamard, B. Cochener, and G. Cazuguel, "Realtime task recognition in cataract surgery videos using adaptive spatiotemporal polynomials," *IEEE Trans Med Imaging*, vol. 34, no. 4, pp. 877–87, Apr 2015.

[12] M. M. Dundar, G. Fung, B. Krishnapuram, and R. B. Rao, "Multipleinstance learning algorithms for computer-aided detection," *IEEE Trans Biomed Eng*, vol. 55, no. 3, pp. 1015–21, Mar. 2008.