IJCTA, 9(7), 2016, pp. 3117-3122 ISSN: 0974-5572

A CAD System for Mammographic Image Analysis using Wavelet Neural Network

Dheeba J.*, J. Amar Pratap Singh and N. Albert Singh*****

ABSTRACT

One of the most leading reason for death among women is breast cancer. As early as the cancer is detected the survival rate will be high. Detecting masses at the early stage using mammograms is the prime and difficult process as most of the cancer masses get obscured in normal breast tissues. This paper presents a new computational methodology that assists the radiologist in detecting masses in mammographic images. The proposed mass detection algorithm is first preprocessed by enhancing the image using histogram equalization and removing the objects surrounding the breast region using global thresholding. The second stage is extraction of shape and texture features for discriminating masses and non-mass regions. The discriminating power of the features extracted is done using classification approach. The third stage involves classification of cancerous tissues using wavelet neural network (WNN). The WNN classifier as evaluated in mammographic images acquired from mammographic analysis centers. Performance analysis of the proposed WNN classifier as based on the accuracy of the classifier to identify pathologies in the mammogram image. A classification accuracy of 90.19% as obtained and the proposed classifier was found to be efficient.

Index Terms: Mammography, Computer Aided Diagnosis, Mass, Mammograms, Wavelet Neural Network, Texture features, Shape features.

1. INTRODUCTION

The second most widespread malignancy among women is breast cancer (BC). The number of women getting affected by BC is increasing every year in western countries [1]. The best possible way to raise the survival rate is detecting the cancer at the earliest. Mammography is the best known technique to analyze the breast at the beginning stage. The technique takes the projects of the breast using x-rays.

A mass in a mammogram have varying margins, different density and diverse in shapes. Masses may be benign/malignant mass. Benign masses normally are smooth with well defined margins and are not shown explicitly. Malignant masses are poorly defined with irregular margins. The breast of younger women is denser than older women and most of the masses get occluded in the normal breast tissue. A purely round and oval shaped mass are likely to be benign and if a mass appears as a lobule then it can be a cancer tissue and should be analysed. Irregular, ill defined and spiculated masses are highly suspicious of malignancy. This shows the importance of analyzing the shape based and texture based features of mammograms in detection process. Finding a breast mass in a mammogram is a tedious task as they are irregular in appearance and has no definite margins [1]. Hence, in this paper we concentrate on detecting malignant masses in mammograms.

Double reading of mammograms by two radiologists can be an effective alternative in finding malignancies. But, this imparts additional manpower which might not be possible in many nations. The studies on mammographic analysis show that radiologists miss visual clues of cancers at a rate of 10%–30% in screening studies [2].

* Associate Professor, Dept. of CSE College of Engineering Perumon, *Email: dheeba.jacob@gmail.com*

** Professor, Dept. of CSE Noorul Islam University, Thuckalay, *Email: japsindia@yahoo.com*

*** SDE (IT), BSNL, Nagercoil, *Email: albertsingh@rediffmail.com*

Computer-aided detection (CAD) system is a more useful tool for doctors in analysing pathologies in breast. The CAD system helps in analysis of interested regions in the breast thereby functions as a second reader tool for doctors. The application of automated tools in screening mammograms is increasing. CAD systems combine methodologies from digital imaging and artificial intelligence along with knowledge from radiologists for making decisions in pathology detection.

Researchers have been developing many methods for early mass detection in mammograms. One among the wide range of research is the CAD of breast cancer [3]. Wener Borges et al. [4] proposed a classifier to classify the mass regions using SVM classifier by performing the classification process in four stages. Many researchers presented an overview of breast abnormalities and methods to detect them [2]. A support vector machine is used to classify breast abnormalities. The classifier based on radial basis function kernel function achieved an overall accuracy of 92.105%. Liyang Wei et al. [5] proposed an automatic detection of microcalcification clusters using a relevance vector machine (RVM). Guido and Nico [6] presented a pixel based classification which uses feedforward neural network for mapping features to the measure of suspiciousness.

In this paper, we present a novel CAD methodology, which uses a WNN to detect and find any abnormalities in the mammograms. The CAD system classifies the pixels as benign/malignant with the help of the texture and shape based features extracted during the feature extraction phase. As the radiologists has to analyze more number of mammograms every day, these computerized tools can be used to assists the radiologists as a second visualization method. The potential of WNN is measured using receiver operating characteristic (ROC) curves. The WNN is trained and tested using mammogram database of 64 mammographic images collected from various screening centers.

2. MATERIALS AND METHODS

The CAD system proposed in this paper mainly possesses 3 stages: pre-processing using histogram equalization and global thresholding, feature extraction using texture and shapes features, classification using WNN classifier. The detailed description is given in the following sections.

2.1. Database of Mammograms and pre-processing

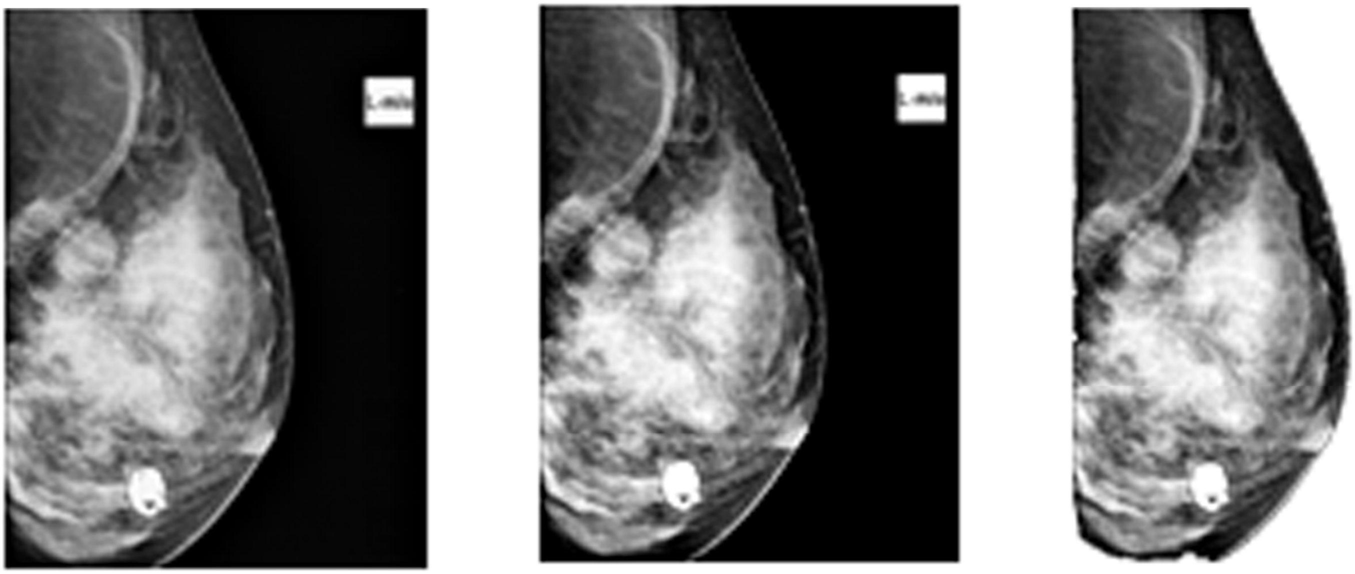
This study used mammograms taken from 64 patients in various screening centers. 256 real clinical mammogram were used for analysis of the proposed methodology. Digital mammograms were obtained using a digitizer, CADPRO advantage. The digital mammogram image was made as a 2020 × 2708 pixel image.

Histogram equalization of the mammogram image is done to enhance the contrast of the mammogram [7]. Mammograms are taken in different illuminations and hence the contrast of the mammogram is changed, so that the histogram, based intensity has a defined shape. The cumulative distribution function (CDF) is found by changing the intensity levels, in a way by stretching the peaks and compressing the throughs. The probability density function f is calculated as in equation (1), where i_k is the intensity, p_k is the number of pixels with intensity level i_k and n is the number of pixels distributed in the image. The CDF calculated using equation (2).

$$f\left(i_k\right) = \frac{p_k}{n} \tag{1}$$

$$F\left(i_k\right) = \sum_{j=0}^{k} f\left(i_k\right) \tag{2}$$

The acquired mammogram and the mammogram obtained after the histogram equalization process is given in Fig. 1.



(a) (b) (c)

Figure 1: (a) Digital mammogram with markers and background
(b) Mammogram after histogram equalization
(c) Region of Interest (Breast region)

The enhanced mammogram contains patients name, their age, the type of projection and background information. By removing these elements from the mammogram image, the power of the classifier is increased by concentrating more on the internal structure of the breast.

The mammogram image contains unnecessary information which may disrupt the detection process. Hence the unnecessary information is segmented and the breast region is separated from the mammogram using global thresholding technique. The image subjected to histogram equalization is given as a n input for global thresholding. In this method pixels that are alike in the mammogram are grouped together. Consider $f(x, y)$ to be the histogram equalized image. A threshold value is fixed; the thresholded image is defined as,

$$G(x,y) = \begin{cases} f, & f(x,y) > GT \\ 0, & f(x,y) \le GT \end{cases} \tag{3}$$

GT is the global threshold value and $GT = 30$ is chosen to remove the background of the mammogram image and segment the breast region at the foreground. Fig. 1 (c) shows the output of breast region segmentation, which is the region of interest (ROI) image.

2.2. Feature Extraction

Feature extraction is a major stage in medical imaging as it describes the content of the image using shape, color and texture features. In most cases, medical images carry less information than color images. The shape based feature like Eccentricity, circularity, compactness is used to describe the shapes of the masses in the mammograms. Eccentricity is defined as the spatial distribution of the structure in the image along its axis. Circularity measures the circular nature of the object and compactness. These measures are described in [8]. These shape features describe the masses that can be characterized by shapes. The circumscribed masses are said to have round shapes and such masses can be described by the shape based features. The other malignant masses having irregular shapes and margins can be classified using the texture based feature extraction methods.

Kenneth Ivan Laws [9] proposed a texture extraction method to extract the texture features from images, these features are named as texture energy measures. The texture features were calculated from the preprocessed mammogram image. The five texture descriptors include edge (E), wave (W), level (L), spot (S) and ripple (R). the convolution kernels defined for L, E, S, W and R are (1, 4, 6, 4, 1), (–1, –2, 0, 2, 1), (–1, 0, 2, 0, –1),

(–1, 2, 0, –2, 1), and (1, –4, 6, –4, 1) respectively. The image obtained from the preprocessing stage is applied to a non linear filter. The image is separated into a 15 × 15 matrix and the absolute mean of all the values were taken. The texture features for the preprocessed image is obtained using the equation (4).

$$TX(x, y) = \sum\sum \left| G(x+i,\ y+j) \right| \tag{4}$$

$TX(x, y)$ were normalized for zero-mean.

2.3. Classification Using WNN

A wavelet neural network based classification system is used to obtain the potential mass regions in the mammograms.

WNN is a proficient model for non linear system analysis [10]. The proposed network combines the strength of wavelet theory and the features of artificial neural network. The WNN is a well known tool that deals with problems in high dimensional model. As such an artificial neural network, the WNN is also comprised on 3 layers. The first layer is the input layer, which takes the features as the input to the network. The second layer is the hidden layer, which is activated using the wavelets as activation function. The network uses a single hidden layer with 120 neurons in that layer. The third layer is the output layer with 2 neurons giving the output of the classification. The output layer neurons are activated using linear activation function.

The network is learned using the backpropagation learning algorithm, which initializes the network parameters with random weights. Assume a value for the learning rate and momentum constant and the network is subjected for learning. The network is learned until the error E_j is acceptably low. The hidden layer of the WNN uses wavelet activation function with dilations and translations, given as λ and t respectively. The output of the WNN uses a linear activation function which is given as,

$$y_j = \sum_i w_{ij} \psi_j(x) + b_j \tag{5}$$

Where w_{ij} is the weight between the hidden and $\psi_j(x)$ output layer and is the wavelet function, a daubechies mother wavelet is used and b_j is the bias. The error E_j is calculated as $(d_j - y_j)$ where d_j is the desired output

Algorithm:

---

for *i* = 1 *to* N
begin
hidden neurons = 120
input neurons = 28, **output neurons** = 2
LR = 0.01,
MC = 0.9
Assign random weights
end
repeat
for each input–output pair in training set
begin
train the patterns

$$E_j = \frac{1}{2} \sum_j \left(d_j - y_j \right)^2 \tag{6}$$

update weights
end
until E_j is acceptably low

---

and y_j is the actual output. The algorithm for training the WNN is outlined below, where MC is the momentum constant and LR is the learning rate used for setting the learning parameters.

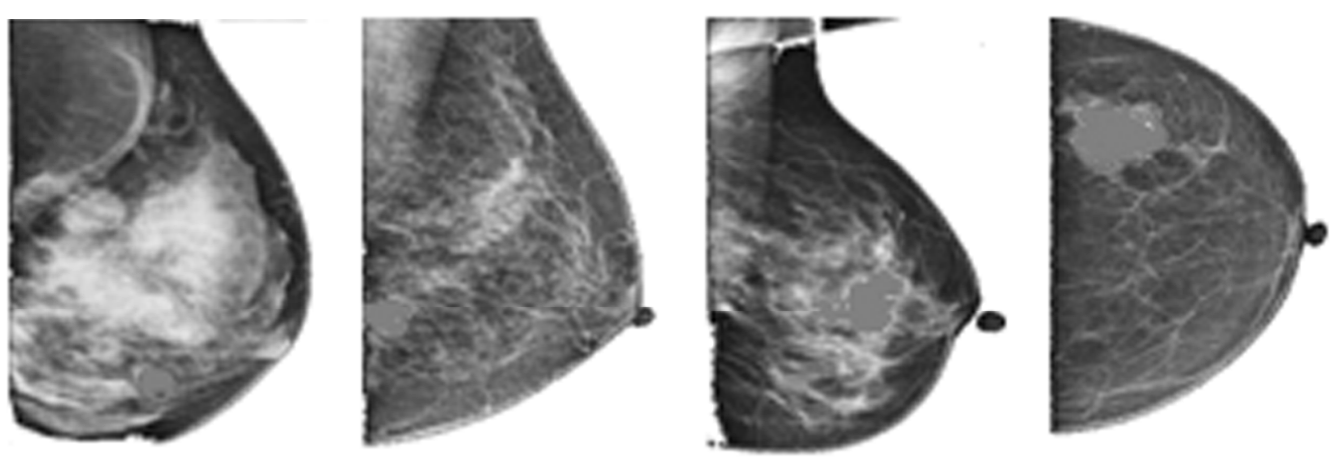
3. PERFORMANCE ANALYSIS

The proposed WNN classifier is trained using the shape and texture features extracted in feature extraction phase. The features include a total of 28 features for each pixel. These are given as input to the classifier and hence the WNN contains 28 input neurons, and one output neuron (benign/malignant mass). The output neuron decides whether the given input is an malignant mass or a benign mass. The hidden neurons is 120 and the activation function used is a wavelet. The output is activated using a linear activation function. The learning rate is set to 0.01 and the momentum constant to 0.9. The training is done until the training error reaches 0.1 or the number of epochs reaches to 500.

The WNN classifier was tested with the mammograms from database obtained in mammogram screening centers. The database contains 216 mammograms of abnormal cases (malignant) and 60 mammograms of normal cases (benign). From these mammograms ROI is chosen and features are extracted, 1064 patterns were taken for training the WNN classifier. Testing was done on the whole set of images in the database. The classified mammogram images identifying the potential mass regions are shown in Fig. 3.

The diagnostic test results yields true positive, false positive, true negative or false negative. To evaluate the performance of the proposed CAD system based on WNN classifier ROC curve is used and the statistical parameter associated with the curve is used to analyse the effectiveness. To plot the ROC curve the sensitivity and specificity should b calculated. In medical terms, sensitivity is the number of images having malignant masses and whose test results are positive for cancer divided by total number of images with malignant masses. Specificity refers to the number of images without malignancy whose test results are negative divided by total number of images without malignancy. If the test is perfect then the test output is 100% successful and hence lies at the upper left corner of the curve. If the test results in a high accuracy then the system is more accurate in discriminating presence and absence of cancer. The neuron in the middle layer depends on size of the training data. The number of input-output patterns used for training of the learning algorithm is 1064 pixels, containing masses, based on this training set the number of neurons in the hidden layer is 120.

The ROC is plotted by varying the neurons of hidden layer and a higher performance is achieved when the hidden neurons reached 120. Table 1 demonstrates the test results carried out in a set of 216 images using WNN classifier. The sensitivity achieved is 95.6% with a specificity of 65.2%.



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

Figure 3: Mass Detection Results (a) Circumscribed mass (b) obscured mass (c) spiculated mass (d) ill defined mass

Table 1
Performance measures of wnn classifier

Metrics	*Performance Measurements*
Sensitivity (%)	95.6%
Specificity (%)	65.2%
Accuracy (%)	90.19%
Misclassification rate	0.1321

From the results tabulated it is evident that the wavelet neural network applied to real clinical database used for classification performed well achieving good classification accuracy. The results carried out shows the strength of WNN classifier that uses a weight updating learning algorithm to learn the network and classify the images containing pathology. The WNN classifier as evaluated in mammographic images acquired from mammographic analysis centers. Performance analysis of the proposed WNN classifier as based on the accuracy of the classifier to identify pathologies in the mammogram image. A classification accuracy of 90.19% as obtained and the proposed classifier was found to be efficient.

4. CONCLUSION

In this paper a novel WNN classifier is used to classify the normal and abnormal tissues in digital mammograms. The classifier is built based on the shape and texture features extracted from the mammogram images. The proposed mass detection algorithm is first preprocessed by enhancing the image using histogram equalization and removing the objects surrounding the breast region using global thresholding. The second stage is extraction of shape and texture features for discriminating masses and non-mass regions. The discriminating power of the features extracted is done using classification approach. The third stage involves classification of cancerous tissues using wavelet neural network (WNN). The accuracy achieved was 90.19% with a misclassification rate of 0.1321.

REFERENCES

[1] American Cancer Society "Cancer Facts and Figures", Atlanta, pp. 1-52, 2015.

[2] Efthimios Motakis, Anna V. Ivshina and Vladimir A. Kuznetsov, "Data-driven approach to predict survival of cancer patients", IEEE Engineering in Medicine and Biology Magazine, vol. 28, no. 4, pp. 58-66, 2009.

[3] Dheeba, J., "Breast cancer diagnosis: an intelligent system using wavelet neural network", AISC, pp. 111-118, 2014.

[4] Wener, Moraes, Aristo, Anselmo Paiva, Marcelo, "Detection of masses in mammogram images using CNN, geostatistic functions and SVM", CBM, vol. 41, 653-664, 2011.

[5] Liyang, Yongyi, Nishikawa, Wernick, and Edwards, "Relevance vector machine for automatic detection of clustered microcalcifications", IEEE Transactions on Medical Imaging, vol. 24, no. 10, pp. 1278-1285, October, 2005.

[6] Guido Te Brake, M., and Karssemeijier, N., "Single and multiscale detection of masses in digital mammograms. IEEE Trans. on medical imaging, vol. 18, pp. 628-639, 1999.

[7] J. A. Stark, "Adaptive image contrast enhancement using generalizations of histogram equalization," IEEE Transactions on Image Processing, vol. 9, no. 5, pp. 889-894, 2000.

[8] Zheng. B., Chang. Y.H. and Gur. D. "Computerized detection of masses in digitized mammograms using single-image segmentation and a multilayer topographic feature analysis", Academic Radiology, vol. 2, no. 11, pp. 959-966, 1995.

[9] K. Laws, Rapid texture identification. In SPIE Vol. 238 Image Processing for Missile Guidance, pp. 376-380, 1980.

[10] Jun Zhang, Walter, G.G., Miao, Y. and Wan Nagi Wayne Lee, "Wavelet neural networks for function learning", IEEE transactions on signal processing, vol. 43, no. 6, pp. 1485-1497, 1995.