

Neural Network Based Classification of Digital Mammograms using DCT Coefficients

Assistant Prof T. Krishna Chaitanya¹, P. Chandra Sekhar Azad²

¹Electronics and communication Engineering, Bapatla Engineering College

²Electronics and communication Engineering, Bapatla Engineering College

Abstract-Breast cancer is very common among the World's women population. Early detection of breast cancer can reduce the effect. Mammography based diagnosis is considered as the most effective to detect breast cancer. But, mammogram images often lead to misdiagnosis due to their low contrast nature. Recent studies show that mammography based diagnosis fails to detect cancerous lumps in the breast in fifteen out of one hundred patients. In this paper, we proposed a neural network based breast cancer detection technique. In the proposed method mammogram images are enhanced and Region of Interest (ROI) extracted to locate the malignant lumps in the women's breast. The Discrete Cosine Transform coefficients of ROI images are obtained, a simple multilayer feed forward neural network is used to train with these coefficients and tested in order to classify the Digital Mammograms. The results are compared with the existing method and the superiority of the proposed method in terms of various metrics is justified.

Keywords- Breast cancer, Region of Interest, Discrete Cosine Transform, Neural Network.

I. INTRODUCTION

Breast cancer is the leading cause of death among females. It is the second reason of deaths of women after lung cancer [1]. Statistics have shown that one out of each ten women is affected by breast cancer. Detection and diagnosis of breast cancer in its early stages increases the chances for successful treatment and complete recovery of the patient [2]. However, although breast cancer incidence has increased over the past decade, breast cancer mortality has declined among women of all ages [2]. This favorable trend in mortality reduction may relate to the widespread adoption of mammography screening, and improvements made in breast cancer treatment.

Breast cancer can be detected by using imaging techniques. Some of the widely used imaging techniques are Magnetic Resonance Imaging (MRI), digital imaging, ultrasound imaging, and mammogram imaging. In MRI radio waves and magnetic fields are used to diagnose the breast cancer. In this method some kind of dye material is injected into the patient's body and reaction of the tumor tissue to this dye material is monitored. The main limitation of the MRI is the cost, which is five times more than that of X-ray mammography. In digital imaging a detector is used to absorb the X-rays in order to form an image. Some of the examples of digital imaging techniques include stereotactic imaging, full field digital mammography, single energy X-ray technique, 3D digital construction, tomosynthesis, and computer aided diagnosis. 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. Mammographic screening has been shown to be effective in reducing breast cancer mortality rates by 30–70% [2].

By using mammography the radiologists detect abnormal masses in the breast. But, there are some limitations of mammography. The quality of mammogram images depends on the density of the breast tissues. Moreover, the mammogram images have low contrast nature. Hence, there are always chances of 'false negative' and 'false positive' results. 'False negative' occurs more often among younger women than among older women, because younger women have dense breast tissues. Breast tissues become somewhat less dense at the old age. 'False-negative' results can delay cancer treatment and promote a false sense of security. On the other hand 'false-positive' mammogram looks abnormal, but no cancer is actually present. 'False-positive' is also more common in younger women because of their dense breast tissues. It is also common among the women who have breast cancer in the family, or who are taking estrogen. 'False-positive' results incur further investigations. In this case diagnostic mammograms, ultrasound, and sometimes MRI or even biopsy are recommended in order to further investigate the presence of a malignant cancer.

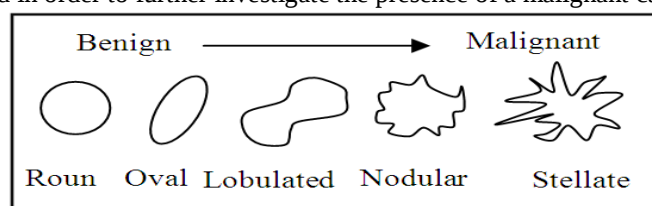


Figure 1. Shapes of breast masses related to cancer (stage wise)

The other limitations of mammography are associated with irregular shapes and locations of cancerous masses in the breast. There are different shapes that have been identified and published in the literature [2] as shown in Figure 1. It is already established that a well-defined round shape masses are considered benign and an irregular shape masses are considered malignant. Sometimes the masses are hidden in the breast tissue and hence they are difficult to trace. In order to overcome these limitations computer aided detection is important. Computer aided detection can help the radiologists to detect and classify the masses as benign or malignant. Computer Aided Detection can also reduce the variable interpretation of the masses by the radiologists.

Breast cancer is the most frequently diagnosed nonskin cancer and the leading cause of cancer-deaths among women. With the advances in digital image processing techniques, it is envisaged that Computer aided diagnosis (CAD) systems can be devised to claim results at par with that of a histopathologist. Image-processing techniques and RBFN can be used to detect malignancy in histopathological images. This gives a wholesome, complete and automated detection of malignancy using both image processing techniques and RBFN [3].

II. REGION OF INTEREST

Breast cancer is still one of main mortality causes in women; but the early detection can increase the chance of cure [1]. Microcalcifications are small size structures, which can indicate the presence of cancer since they are often associated to the most different types of breast tumors. However, they very small size and the X-ray systems limitations lead to constraints to the adequate visualization of such structures, which means that the microcalcifications can be missed many times in mammogram visual examination. In addition, the human eyes are not able to distinguish minimal tonality differences which can be another constraint when mammogram image presents poor contrast between microcalcifications and the tissues around them. Computer-aided diagnosis (CAD) schemes are being developed in order to increase the probabilities of early detection [5].

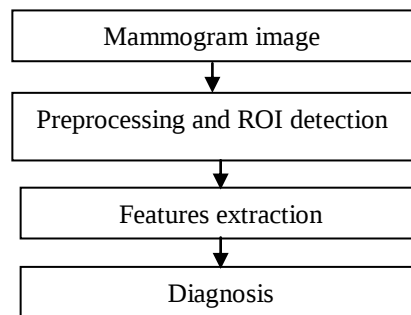


Figure 2.*The CAD system architecture to extract ROI*

A general CAD system can be divided into three parts as shown in Fig.1. The first part is to find region of interest (ROI) which have high possibilities of breast cancer. Finding ROIs with high accuracy can be very important for correct diagnosis. The second part is to extract appropriate features from ROIs for the next part. And the last part of a CAD system is to determine whether a region contains malignant symptoms or not based on the feature values. Effective identification of the masses in breasts is based on some important features like textural, intensity, and shape extracted from the mammographic images. These features have been used for detecting masses in the breast.

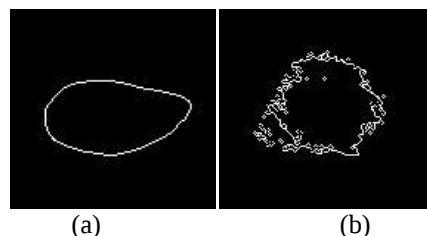


Figure 3.*Examples for ROI are of (a) benign mass and (b) malignant masses.*

III. NEURAL NETWORKS

The simplest definition of a neural network, more properly referred to as an 'artificial' neural network (ANN), is provided by the inventor of one of the first neurocomputers, Dr. Robert Hecht-Nielsen. He defines a neural network as a computing system made up of a number of simple, highly interconnected processing elements, which process information by their dynamic state response to external inputs.

ANNs are processing devices that are loosely modeled after the neuronal structure of the mamalian cerebral cortex. A large ANN might have hundreds or thousands of processor units, whereas a mamalian brain has billions of neurons with a corresponding increase in magnitude of their overall interaction and emergent behavior. Although ANN researchers are generally not concerned with whether their networks accurately resemble biological system, but they compared it like the function of the retina and modeled it as the functionality of an eye.

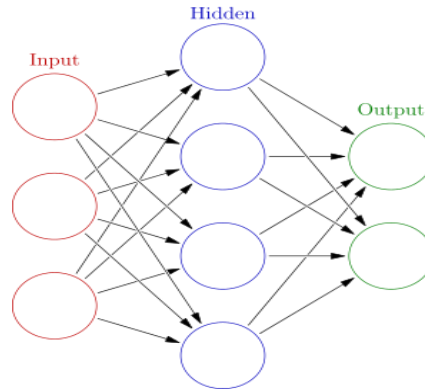


Figure 4. A fully connected multilayer feedforward network with one hidden layer

Neural networks are typically organized in layers. Layers are made up of a number of interconnected 'nodes' which contain an 'activation function'. Patterns are presented to the network via the 'input layer', which communicates to one or more 'hidden layers' via a system of weighted 'connections' 'will perform the actual process. The hidden layers then link to an 'output layer' to generate output. Figure 4 shows a multilayer feedforward ANN where all the neurons in each layer are connected to all the neurons in the next layer with some weight, which is called as a fully connected network.

IV. PROPOSED METHOD

In this work we consider three sample mammogram images namely (a) normal breast, (b) breast with benign masses, and (c) breast with malignant masses shown in figures.

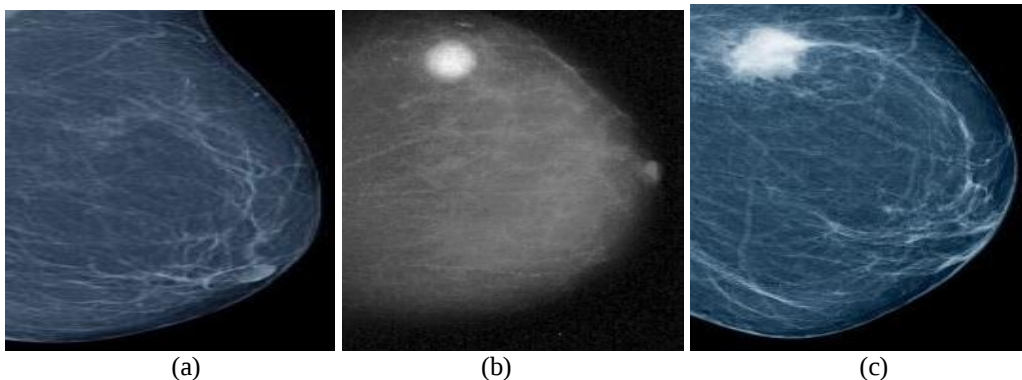


Figure 5. Mammogram images of (a) normal breast, (b) breast with benign mass, (c) breast with malignant Masses

In the proposed method the detection will be done as follows,

- i. Image enhancement to separate masses.
- ii. ROI extraction.
- iii. Calculate DCT coefficients of ROI's.
- iv. Training and testing using neural networks.

4.1. Image enhancement to separate masses

The mammogram images need to be enhanced so that the masses can be separated from the rest of the mammogram image. In order to enhance mammogram image we follow the steps shown in Figure 7. In the first step we pre-process the image and convert the image into a binary image. The other major steps are converting the color image into a grey image, filtering the image by Gaussian filter, and applying a threshold to separate the masses from the rest of the breast image [6].

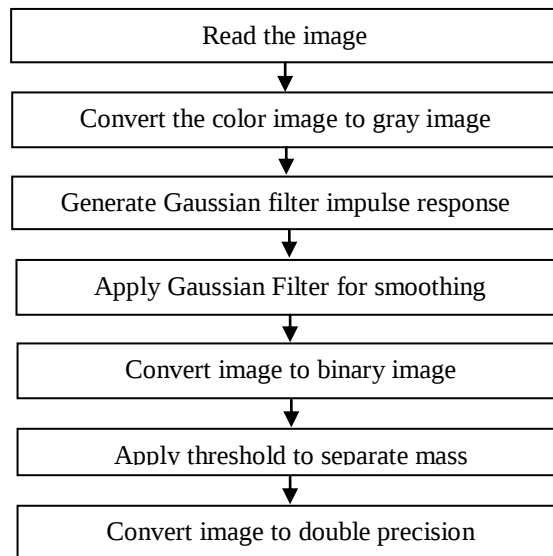


Figure 6. Steps to separate the masses from images

4.2. ROI Extraction

The next step followed by image enhancing stage is, ROI extraction in order to get the tumor area, we need to detect the edges and we will take region of interest (ROI) as mentioned in previous chapter.

4.3. DCT Coefficients of ROI Images

Assume that the data array has finite rectangular support on $[0, N_1 - 1] \times [0, N_2 - 1]$, then the 2-D DCT on an image can be given as

$$X_C(k_1, k_2) \triangleq \sum_{n_1=0}^{N_1-1} \sum_{n_2=0}^{N_2-1} 4x(n_1, n_2) \cos \frac{\pi k_1}{2N_1} (2n_1 + 1) \cos \frac{\pi k_2}{2N_2} (2n_2 + 1), \quad \rightarrow (1)$$

For $(k_1, k_2) \in [0, N_1 - 1] \times [0, N_2 - 1]$; Otherwise, $X_C(k_1, k_2) \triangleq 0$.

The DCT basis functions for size 8×8 , the mapping between the mathematical values and the colors (gray levels) is the same as in the DFT case. Each basis function occupies a small square; the squares are then arranged into as 8×8 mosaic. Note that unlike the DFT, where the highest frequencies occur near $(N_1/2, N_2/2)$, the highest frequencies of the DCT occur at the highest indices $(k_1, k_2) = (7, 7)$.

4.4. Training and Testing Using Neural Networks

In the final stage, classification of coefficients is performed by using neural networks [7]. The DCT coefficients of ROIs are tested against the already trained coefficients of reference ROI templates. The obtained coefficients of ROI's are given as test inputs to the ANN, in order to determine whether it is Benign or Malignant tumor [8].

V. EXPERIMENTAL WORK & RESULTS

For testing the proposed work the mammogram data base set MIAS of 322 images was chosen. The images were processed to extract the ROI, and these ROI images are processed to extract the DCT coefficients from them. A multilayer feed forward Neural network was chosen for training the DCT coefficients initially with 10-fold cross-validation, later the trained neural network will be used for test whether the image is Benign or Malignant. Further the work was extended for different cross-validations to ensure the superiority of the algorithm. The performance measure of the classifiers in detecting the breast cancer can be evaluated using the parameters like Precision, Recall & Accuracy.

Accuracy is the percentage measure of correctly classified instances for all instances. It can be obtained as below.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad \rightarrow (2)$$

Precision is of correctly classified instances for those instances that are classified as positive, and it is calculated using the equation.

$$\text{Precision} = \frac{TP}{TP+FP} \rightarrow (3)$$

Recall is the measure of the positive instance that is correctly classified, and it can be calculated with below equation.

$$\text{Recall} = \frac{TP}{TP+FN} \rightarrow (4)$$

Where TP : True Positive
FP : False Positive
TN : True Negative
FN : False Negative

Table 1. Comparison with existing method for 10-fold cross-validation

	Accuracy	Precision	Recall
RBF	73%	0.72	0.72
Proposed NN-DCT	91.81%	0.9466	0.88

Table 2. Validity assessment by using proposed method for different cross-validations

	Accuracy	Precision	Recall
10 - fold	91.81%	0.9466	0.88
20 - fold	85%	0.8796	0.8016
30 - fold	80.76%	0.879	0.706

The results mentioned in Table 2 are the average of the several iterations by varying the training and testing data set, by randomly choosing the images form MIAS database with the mentioned cross validations.

VI. CONCLUSION

This paper has attempted to develop new technique for mammograms classification. Although mammogram images are used for cancer screening, it has some limitations due to its low contrast nature. Moreover, early detection of breast cancer is not straight forward. The Classification of the mammogram images with a trained neural network having DCT coefficients of the ROI images as input will generate a better accuracy. We also show that our method produces better average accuracy even for less training data set. By combining ROI of images, DCT coefficients, & neural networks it has been shown that our proposed system can improve the average classification rate of mammograms.

Acknowledgement

This work was supported by University Grants Commission, India under the funding scheme – Minor Research Project. The authors thank UGC for their support towards research under this scheme.

VII. REFERENCES

- [1] Sandra Buseman, M.D., M.S.P.H., Judy Mouchawar, M.D., M.S.P.H. ,Ned Calonge, M.D., M.P.H. ,Tim Byers, M.D., M.P.H. “Mammography Screening Matters for Young Women with Breast Carcinoma”.
- [2] Huda Al-Ghaib, Reza Adhami, and Melanie Scott, “An overview of mammogram analysis”, 0278-6648/16, 2016 IEEE.
- [3] Mahendra G. Kanojia, Siby Abraham, ” Breast Cancer Detection Using RBF Neural Network”, 978-1-5090-5256-1/16/\$31.00_c 2016 IEEE.
- [4] J.C. Fua,, S.K. Leeb, S.T.C. Wongc, J.Y. Yeha, A.H. Wanga, H.K. Wud, ”Image segmentation feature selection and pattern classification for mammographic microcalcifications”, J.C. Fu et al. / Computerized Medical Imaging and Graphics 29 (2005) 419–429.
- [5] JuCheng Yang, DongSun Park, “Detecting Region-of-Interest (ROI) in Digital Mammogram by using Morphological Bandpass Filter”, 0-7803-8603-5/04/\$20.00200 04 IEEE.

- [6] Fatmah ElZahra, Ameena Hateem, Mussab Mohammad, and Mohammed Tarique, "Fourier Transform Based Early Detection of Breast Cancer by Mammogram Image Processing; JOURNAL OF BIOMEDICAL ENGINEERING AND MEDICAL IMAGING" Volume 2, Issue 4, August, 2015.
- [7] Moh'd Rasoul A, Al-Hadidi, Mohammed Y. Al Gawagzeh and Bayan A. Alsaaidah "Solving mammography problems of breast cancer detection using artificial neural networks and image processing techniques", Indian Journal of Science and Technology- Vol. 5 No. 4 (Apr 2012).
- [8] F. Schnorrenber, N Tsopafoulis, S. Collios, M. Vossiliou, A. Adamou, K. Kyriacou. "Improved Detection of Breast Cancer Nuclei Using Modular Neural Networks". IEEE ENGINEERING IN MEDICINE AND BIOLOGY 0739.51 75/00/510.00@2000IEEE.
- [9] Suckling J., et al. Mammographic Image Analysis Society (MIAS) database.
<https://www.repository.cam.ac.uk/handle/1810/250394>.