Breast Cancer Detection using Two Dimensional Principal Component Analysis and Back Propagation Neural Network

Dr. A. Venkataramana
Faculty in ECE, Q.Q. Government Polytechnic,
Department of Technical Education,
Hyderabad, Telangana, India, 500064
Email: venkat_ramana33@yahoo.com

G. Prashanth
M.Tech. (ECE),
JNT University, Hyderabad,
Telangana, India
Email: prashanthganji07@gmail.com

G. Mounika
B.Tech 4th year EEE,
Swami Ramananda Tirtha Institute of Science & Technology, Nalgonda,
JNTUH, Telangana, India
Email: ganjimounika89@gmail.com

R. Vinay Krishna
B.Tech, 3rd year ECE,
JNTU, Jagtial,
Telangana, India
Email: vinaykrishnaraghu100@gmail.com

Abstract— Breast cancer is the most common cancer which affects women around the world. It has been increasing over the years. Detection and diagnosis is the main factor for breast cancer control which increases the success rate of treatment, saves lives and reduce the cost. This paper proposes an efficient approach for breast cancer detection in mammogram breast images using two dimensional principal component analysis and back propagation neural network. The proposed approach consists of four step by step procedures namely preprocessing of breast images, image enhancement, feature extraction and classification. Two dimensional principal component analysis is used to obtain the features of the preprocessed and enhanced image. The reason for selecting two dimensional principal component analysis is it is easier to evaluate the covariance matrix accurately and less time is required to determine the corresponding features. Finally, Back propagation neural network is used to classify whether the given mammogram image is normal or abnormal. Simulation results are carried out using the proposed approach by considering MIAS data base. From the results, it is observed that proposed approach provide better accuracy.

Keywords: *Mammogram images, Breast Cancer Detection, Two Dimensional Principal Component Analysis, Back Propagation Neural Network*

*****

I. INTRODUCTION

Breast can be classified into two types namely fatty breast and glandular breast. When the amount of fat
tissues exceed the amount of fibro glandular tissues, the breast can be classified as fatty breast, where as the amount of fibro glandular tissues exceed the amount of fatty tissues, the breast can be classified as glandular breast. Breast cancer occurs when breast tissues grow, change and multiply rapidly without control which may form lump or mass of extra tissues. These masses are called tumor and can be non-cancerous (benign) or cancerous (malignant) [1]. Breast cancer is the most common cancer which affects women around the world. It has been increasing over the years. Breast cancer can be divided into four different stages like micro-classification, masses, architectural distortion and bilateral asymmetry. Breast cancer not only affects women, it can also be present in some man. Hence, screening need to be done for both. But, highest percentage of cancer get affected in women so the world is mostly addressing the women health care. Successful treatment of breast cancer depends on early detection.

Some of the image techniques used to capture the breast of suspected person are magnetic resonance imaging (MRI), ultrasound and digital mammograms. Among the imaging techniques, the digital mammogram is inspected to improve the specificity and sensitivity of breast cancer. Further, X-ray mammogram is the most effective and economical breast imaging modality due to its simplicity, portability and cost effectiveness. Mammography is used as a breast examination where the patient is undressed to the waist and stand infront of the X-ray machine. Then, each breast is compressed between two plates to take images of each breast by using a brief X-ray pulse [2].

During the last few years, various methods and techniques are applied for image processing, tumor detection and tumor classification. Fatima Eddaoudi et al. [4] proposed mass detection threshold and classified using SVM classifier. In their work,
features are calculated from the co-occurrence matrix. The test of these methods on mammograms of MIAS databases showed better performance in detecting masses compared to the methods proposed in the literature. Tingting Mu et al. [5] proposed a new approach using strict two-surface proximal (S2SP) classifier for tumor classification which uses 22 features of the segmented tumor portion. The earlier detection of breast cancer using self-similar fractal method [6] was proposed by BhagwatiCharan Patel et al. The method was tested over several images of image databases taken from BSR APPOLO for cancer research and diagnosis, India. Olfati. E et al. [7] presented eigen factors for comparison and classification and the results were compared with GA based results. A swarm intelligence technique based support vector machine classifier (PSO_SVM) is proposed for breast cancer diagnosis in [8]. Mammography feature analysis and mass detection in breast cancer images was presented by by Patel et al. [9]. Karthikeyan Ganesan et al. [3]] studied various methods, issues and challenges of computer aided diagnosis. Jinshan Tang et al. [3] developed a CAD system for cancer detection and diagnosis. The techniques used are applied for various stages like finding, locating, detection and classification. Mohammad Rabab et al. [3] applied feature extraction method which was used to check the tumor is benign or malignant. The literature survey says that there are various methods used for analysis of medical images. The various classification techniques applied are classifying the images with less number of features. Due to less number of features, the classification accuracy is also less and it is limited on the type of input images. The ability to improve diagnostic information from medical images can further be enhanced by designing computer processing algorithms, application and software. Hence, an efficient technique to detect cancer in mammogram breast images using two dimensional principal component analysis (2D-PCA) and back propagation neural network is proposed in this paper. 2DPCA is used for extraction the features of breast images. The reason for selecting 2D PCA [10] is as opposed to conventional PCA, 2DPCA is based on 2D image matrices rather than 1D vectors so the image matrix does not need to be transformed into a vector prior to feature extraction. Instead, an image covariance matrix is constructed directly using the original image matrices, and its eigenvectors are derived for image feature extraction. Further, 2DPCA has two important advantages over PCA namely it is easier to evaluate the covariance matrix accurately and less time is required to determine the corresponding eigenvectors. Back propagation neural network is used to classify whether the given mammogram image is normal or abnormal. The performance of back propagation neural network is better than that of linear classifier. Hence, back propagation neural network is used as classifier in this paper.

This paper is organized into four sections. Section II presents the proposed approach for breast cancer detection in mammogram breast images. The proposed approach uses 2DPCA for feature extraction and back propagation neural network for classification. Hence, the details about 2DPCA and back propagation neural network is also presented in this section. Simulation results are reported in section III. Finally, the last section presents the conclusions about the work.

II. PROPOSED APPROACH

The proposed approach for computer aided diagnosis system for breast cancer detection consists of four step by step procedures namely preprocessing of breast images, image enhancement, feature extraction and classification.

Preprocessing: the purpose of preprocessing is to remove background noise. In this work, we used median filtering for smoothing. The median filter is a non linear filter which removes speckle noise and salt and pepper noise in the image while preserving the edges of the image. In median filtering operation, the pixel values of neighbourhood window are ranked according to the intensity and the middle value becomes output value for the pixel.

Image enhancement: the median filtered image is then subjected to image enhancement. The purpose of image enhancement is to process an image so that the result is more suitable than the original image for specific application. In this work, histogram equalization technique is used for image enhancement since it is simple and effective image enhancement technique for image contrast enhancement. This technique has very less complexity and produces rapid computation compared to other techniques.

Feature extraction: Feature extraction and selection are important steps in breast cancer detection and classification. An optimum feature set should have effective and discriminating features. In this work, we used 2DPCA for feature exraction. Principal component analysis (PCA), also known as KarhunenLoeve expansion, is a classical feature extraction and data representation technique widely used in the areas of pattern recognition and computer vision. In the PCA-based face recognition technique [10] , the 2D face image matrices must be previously transformed into 1D image vectors. The resulting image vectors of faces usually lead to a high dimensional image vector space, where it is difficult to evaluate the covariance matrix accurately due to its large size and the relatively small number of training samples. In the

paper [10], a straightforward image projection technique, called two-dimensional principal component analysis (2DPCA), is developed for image feature extraction. As compared to conventional PCA, 2DPCA is based on 2D matrices rather than 1D vectors. That is, the image matrix does not need to be previously transformed into a vector. Instead, an image covariance matrix can be constructed directly using the original image matrices. In contrast to the covariance matrix of PCA, the size of the image covariance matrix using 2DPCA is much smaller. As a result, 2DPCA has two important advantages over PCA namely it is easier to evaluate the covariance matrix accurately and less time is required to determine the corresponding eigenvectors. Hence, 2DPCA is used for extracting features of breast images in this paper.

Let X denotes an n-dimensional unitary column vector. The idea presented in [10] is to project image A, an m*n random matrix, on to X by the following linear transformation Y=AX Thus, an n-dimensional projected vector Y, which is called the projected feature vector of image A is obtained. The total scatter of the projected samples can be introduced to measure the discriminator power of the projection vector X. The total scatter of the projected samples can be characterized by the trace of covariance matrix of the projected feature vectors. The following criteria [10] is used to obtain a good projection.

$J(X) = tr(Sx)$ where Sx denotes the covariance matrix and tr (Sx) denotes the trace of Sx. The physical significance of maximizing the criteria is to find a projection direction X, on to which all samples are projected, so that the total scatter of the resulting projected samples is maximized. The covariance matrix Sx can be denoted by

$Sx = E[(A-Ua)X*(A-Ua)^{T}X]$

Let the image covariance (scatter) matrix be

$Gt = E[(A-Ua)*(A-Ua)^{T}]$. We can evaluate Gt directly using the training image samples. Suppose that there are M training image samples in total, the jth training image is denoted by an n*n matrix Aj (j=1,2….,M), and the average image of all training samples is denoted by Ua. Then, Gt can be evaluated by

$$Gt = \frac{1}{M} \sum [(A-Ua)*(A-Ua)^{T}]$$

The optimal projection axes, X1,……., Xd, are the ortho normal eigen vectors of Gt corresponding to the first d largest eigen values. The optimal projection vectors of 2DPCA, X1,……., Xd, are used for feature extraction. For a given image sample A, then Yk=AXk, k=1,2,……,d. Hence, we obtain a family of projected feature vectors, Y1,……., Yd, which are called the principal component(vectors) of the sample image A. The matrix B = [Y1,……., Yd] is called the feature matrix. Thus, we obtain features of the given mammogram images.

Classification: The extracted features are then applied for classification. The obtained features are applied to back propagation neural network for classification. Artificial neural networks are the collection of mathematical models that imitate the properties of biological nervous system and the functions of adaptive biological learning [3]. It is a self-learning system that changes its parameters based on external or internal information that flows through the network during the learning phase. ANN is composed of an input layer, an output layer and one or more hidden layers. Layer is composed of neurons. Back-propagation neural network Back-propagation (BP) neural network is a feed-forward ANN with supervised learning process. Frequently used back-propagation neural networks have one or two hidden layers and 2–10 neurons in each layer. There is no universal rule to decide the number of layers or number of neurons in each layer. The performance of back-propagation neural network is better than that of linear classifiers. Hence, back propagation neural network is used in the proposed approach.

III. SIMULATION RESULTS

In order to verify the applicability of the proposed approach, we have selected mini- MIAS database [11]. The database is freely available for scientific and research purposes. The images of database originate from a film screen mammographic imaging process in the United Kingdom National Breast Screening program. The algorithm is implemented on 100 images consists of 50 normal and 50 abnormal. Some sample images of database are shown in Fig.1. The size of all images is 1024 x 1024. We used 50% images for training and the remaining 50% for testing. Matlab software is used to implement the proposed technique since Matlab is a high performance language for education and research as it integrates computation, visualization programming. Out of 100 images, 98 images are correctly classified. Hence, the accuracy of the proposed method is 98%.

IV. CONCLUSIONS

This paper proposes an efficient technique for breast cancer detection using two dimensional principal component analysis and back propagation neural network. The proposed approach consists of four step by step procedures namely preprocessing of breast images, image enhancement, feature extraction and classification. In the preprocessing stage, the background noise present in image is removed by median filtering. Histogram equalization is used for image enhancement since it is simple and effective image enhancement technique which produces rapid results compared to other techniques. Two dimensional principal component analysis is used to obtain the features of the preprocessed and enhanced image. The reason for selecting two dimensional principal component analysis is it is easier to evaluate the covariance matrix accurately and less time is

required to determine the corresponding features. Finally, Back propagation neural network is used to classify whether the given mammogram image is normal or abnormal. It is observed from the simulation results that the proposed approach provide better accuracy.

Our future work will be towards early detection of breast cancer. Our future will also be directed towards automatic detection of tumor in MRI images. I am also planning to connect X-ray camera to the USB port of the computer system which enables the female patient to capture the image of her breast by herself and the send it to a computer which has shared folder. We can make a connection with shared folder and this algorithm. Once the image arrives to the shared folder, the algorithm will run and give a medical report.

Fig.1. Some sample images of MIAS database.
Column1: original images
Column 2: median filtered images
Column 3: Histogram equalized images

REFERENCES

[1] Monica Jenefer and V. Cyrilraj, 2014, " An Efficient Image Processing Method for Mammogram Breast Cancer Detection", Journal of Theoretical and Applied Information Technology, vol.69, no.1, pp. 32-38

[2] Aarthy and Prabhu, 2016, " A Computerized Approach on Breast Cancer Detection and Classification", IIOAB Journal, special issue, vol.7, pp.157-169

[3] HD.Cheng et. Al. , 2010, "Automated Breast Cancer Detection and Classification using Ultrasound Images: A Survey", pattern recognition 43 (2010) pp. 299-317

[4] Fatima Eddaoudi, FakhitaRegragui, AbdelhakMahmoudi, NajibLamouri, 2011 , "Masses Detection Using SVM Classifier Based on Textures Analysis",Applied Mathematical Sciences, Vol. 5, 2011, no. 8, 367 – 379

[5] Tingting Mu,1 Asoke K. Nandi,1 and Rangaraj M. Rangayyan2, Classification of breast Masses using Selected shapes, edge sharpness, teture features with linear and kernalnased classifier, Journal of Digital Imaging, Vol 21, No 2 (June), 2008: pp 153Y169

[6] Bhagwati Charan Patel , Dr. G.R.Sinha , Early Detection of Breast Cancer using Self Similar Fractal Method, International Journal of Computer Applications (0975 – 8887) Volume 10– N.4, November 2010

[7] Olfati, E. , Zarabadipour, H. ; Shoorehdeli, M.A. Feature subset selection and parameters optimization for support vector machine in breast cancer diagnosis , IEEE published , 2014 Iranian Conference on Intelligent Systems (ICIS)

[8] Hui-Ling Chen, Bo Yang, Gang Wang, SuJing Wang, JieLiu,Da You Liu , Support vector machine based diagnostic system for breast cancer using swarm intelligence, Journal of Medical Systems. 2012 Aug;36(4):2505-19.

[9] Vishnu Kumar K Patel et al. , 2012, " Mammogram of Breast Cancer Detection based using Image Enhancement Algorithm", International journal of emerging technology and advanced engineering, vol.2, issue 8, pp.143-147

[10] Jian Yang, David Zhang, Alejandro F. Frangi, and Jing-yu Yang, "Two-Dimensional PCA: A New Approach to Appearance-Based Face Representation and Recognition", 2004, IEEE Transaction on pattern Analysis and Machine Intelligence, Vol. 26, No. 1, 2004, pp. 131-137

[11] J. Suckling et al., (1994) " The Mammographic Image Analysis Society Digital Mammogram Database", Expert Medica International Congress Series 1069, pp. 375-378