ANN based Classifier System for Digital Mammographic Images

Bikesh Kumar Singh [1], Dr. Anamika Yadav [2], Shailaja Singh [3]
[1]Assistant Professor, Biomedical Department, NIT Raipur
[2]Assistant Professor, Electrical Department, NIT Raipur
[3]M.Tech Scholar, NIT Raipur

ABSTRACT
Breast cancer is one of the most common cancers among women of the developing countries in the world, and it has also become a major cause of death [1, 2]. Treatment of breast cancer is effective only if it is detected at an early stage. X-ray Mammography is the most effective technique used by radiologists in the screening and diagnosis of breast cancer in women but the mammographic images are complex [2]. With the development in Artificial Intelligence (AI) and Soft Computing Techniques, Computer-Aided Diagnosis (CAD) attracts more and more attention for brain tumor diagnosis. Computer-Aided Diagnosis system (CAD) can be very helpful in detecting and diagnosing breast abnormalities earlier and faster than typical screening programs. This paper presents retrieval and ANN (Artificial neural network) based classification system for computer aided diagnosis of breast cancer using texture features. The proposed system uses Euclidean distance for the comparison of the feature vector of the query image and each image in the database. It has been found that the proposed CBIR system is gives 80% retrieval accuracy for the database of 200 images of mini-MIAS database. Further the ANN based classifier gives 94% accuracy in classifying benign and malignant breast masses. MATLAB ® 7.01 image processing toolbox and ANN toolbox have been used to implement the algorithm. The results show that texture features can be effectively used for classifying mammographic images with high level of accuracy.

General Terms
Pattern Recognition, Medical Image Processing.

Keywords
Mammograms, image processing, shape and texture features, Content Bases Image Retrieval (CBIR), ANN (Artificial Neural Network).

1. INTRODUCTION
Breast cancer is the most common form of cancer in women. It is responsible for 25% of death in women aged 35 to 54 years [1]**.** Early screening of breast tumor is of great help in the prognosis of a favorable outcome. Mammography is the process of using low-energy-X-rays to examine the human breast and is used as a diagnostic and a screening tool. The detection of micro calcification has been explored by various groups of researchers. Small (sub 15mm), low contrast masses are considered more critical than microcalcifications, since they are more difficult to detect [3]. Breast cancer causes a desmoplastic reaction in breast tissue. A mass is observed as a bright, hyper-dense object. Breast cancer is characterized with the presence of a mass accompanied or not accompanied by calcifications [5].Radiologists need help in identifying difficult to see mass lesion cancers to decrease the number of cancers missed and to reduce the number of unnecessary biopsies of benign tissue. Mammograms are noisy and inconsistent to interpret in their raw form. Thus to produce a reliable representation of the breast anatomy we need to pre-process the mammograms. This paper comprises of studying several enhancement techniques and their performance evaluation on the basis of some parameters like PSNR (Peak Signal to Noise Ratio) and CNR (Contrast to Noise Ratio) [3]. Before feature extraction each image in the database was first enhanced using best enhancement technique [3]. Then a feature vector consisting of texture feature is generated for each image in the database and stored in .mat file in MATLAB environment. Finally ANN based classifier using back propagation algorithm was used to classify whether the mammogram images were having benign or malignant breast tumor. The ANN based classifier system produces 94 percent accuracy. Our database consists of 200 mammograms collected from mini-MIAS database [4].

2. FEATURE EXTRACTION
For each image, we compute number of significant features. Medical images possess different texture depending upon area of body considered for imaging. According to Smith and Chang texture refers to visual patterns which have properties of homogeneity and cannot result from the presence of only a single color or intensity. Texture perception has a very important aspect in the human visual system of recognition and interpretation. Generally speaking, texture feature extraction methods can be classified into three major categories namely: Statistical, Structural, and Spectral [6, 7]. We use Haralick texture features which includes 13 texture features namely Energy, Inertia, Entropy, Inverse Difference Moment, Sum Average, Sum Variance, Sum Entropy, Difference Average, Sum Entropy, Difference Average, Difference Variance, Difference Entropy, Information measure of correlation1, Information measure of correlation2 [13,14]. In order to improve the biopsy yield ratio, masses must be classified as benign or malignant. Masses with radiopaque and with irregular in shapes are usually malignant, and those combined with radiolucent shapes are benign [15].

3. CONTENT BASED IMAGE RETRIEVAL
Content based image retrieval system allows the retrieval of relevant images based on a pre-defined similarity measure between image features. In terms of medical thermal imaging, images that are similar to a sample exhibiting symptoms of a certain disease or other disorder will be likely to show the same or similar manifestations of the disease. These known cases together with their medical reports will then provide a

valuable asset for the diagnoses of the unknown case. Retrieval is often performed in a query by example fashion where a query image is provided by the user [6-10]. The image database is then searching through all images in order to find those with the most similar indices which are returned as the images most alike to the query image. A large variety of features have been proposed in the CBIR literature [7, 12]. In our test database we have total of 200 images consisting of normal and abnormal cases from that 160 images are retrieved correctly, using the Euclidean distance.

3.1 Distance Calculation using Euclidean Distance

In mathematics, the Euclidean distance or Euclidean metric is the ordinary distance between two points. The Euclidean distance between points p and q, if p = (p_1, p_2,..., p_n) and q = (q_1, q_2,..., q_n), then the distance from p to q, or from q to p is given by:

$$d(p,q) = d(q,p) = \sqrt{\sum_{i=1}^{n}(q_i - p_i)^2} \quad (1)$$

3.2 Result of CBIR System

When query image is presented to the CBIR system, it is first enhanced and then a feature vector consisting of texture features was generated and compared with the similar feature vector of each image in the database using Euclidean distance in order to retrieve each most similar image from the database. If the Euclidean distance between the features vector of candidate image i.e the image from database and the query image is zero or below the threshold value then the matched candidate image is shown at the output. In our Test database of 200 images 160 images were retrieved correctly at threshold of 0.1, thereby giving 80% accuracy. This work may assist radiologist to classify cancerous and noncancerous mammograms. Figure1 shows the result of the CBIR system, where left side image is query image and system retrieved 6 matches related to the query image with increasing order of Euclidean distance. Thus the 1st image is the best match to the query image; because it corresponds to minimum Euclidean distance.

4. ANN BASED CLASSIFICATION

Neural networks are those information processing systems, which are constructed and implemented to model the human brain. An artificial neural network is an efficient information processing system which resembles in characteristics with a biological neural network. ANNs possess large number of highly interconnected processing elements called nodes or units or neurons, which usually operate in parallel and are configured in regular architectures. Each neuron is connected with the other by a connection link. Each connection link is associated with weights which contain information about input signal. The information is used by the neuron net to solve a particular problem. ANNs collective behavior is characterized by their ability to learn recall and generalize training patterns or data similar to that of a human brain. They have the capability to model networks of original neurons s found in the brain.

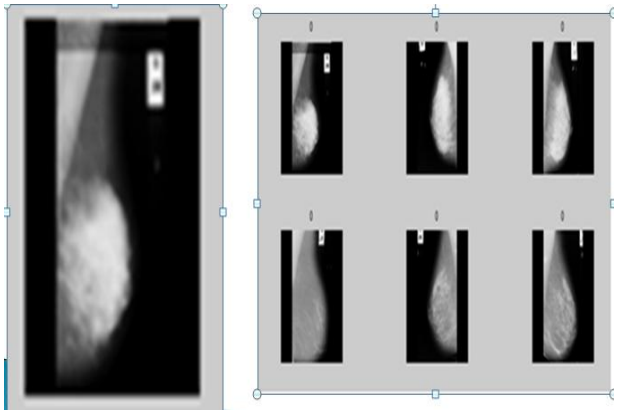

Fig 1: Result of CBIR system

The use of artificial neural networks (ANNs) offers some useful properties and capabilities. The capability to discriminate between samples that are not linearly separable is an important property, particularly if the underlying physical mechanism responsible for generation of the input information is inherently nonlinear [16]. We present results of classification of masses and tumors as benign or malignant.

An artificial neural network is developed with a systematic step-by-step procedure which optimizes a criterion commonly known as the learning rule. The input/output training data is fundamental for these networks as it conveys the information which is necessary to discover the optimal operating point. In addition, a non linear nature makes neural network processing elements a very flexible system. There are usually three phases for detection and diagnosis using neural networks. Phase-1 comprises of the feature extracted data. Phase-2 is preparation of the data pertaining to inputs and target (input-output pairs) which is used for training the network. The network is iteratively adjusted to memorize the input output (target) relationships. In phase 3 the testing and validation process has been carried out

The proposed ANN based Computer Aided Diagnosis system is basically an application of artificial neural network which is used to diagnose the breast cancer in the mammographic images.

4.1 ANN architecture for the classification

The Basic architecture of ANN for the breast cancer tumor classification includes input module, hidden layer and output layer. The texture features were used as an input to the ANN based classifier.

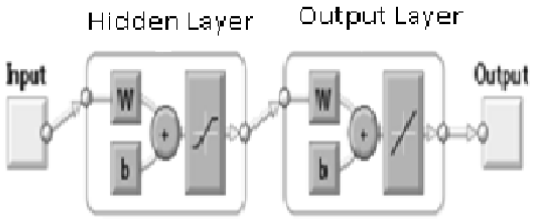

Fig 2: Architecture of Neural Network

The network has 13 normalized inputs and only single output. The number of neurons selected for the hidden layer is 4 as shown in Figure 2 ,which are obtained by trial and error for a given mean square error (MSE), initially for 1, 2, 3,then 4 neurons minimum error is obtained to be as 9.57141e-006 as

shown in figure 3. We experimented on many transfer functions like purelin, logsig and tansig. However, best result was found with hyperbolic tangent sigmoid function.

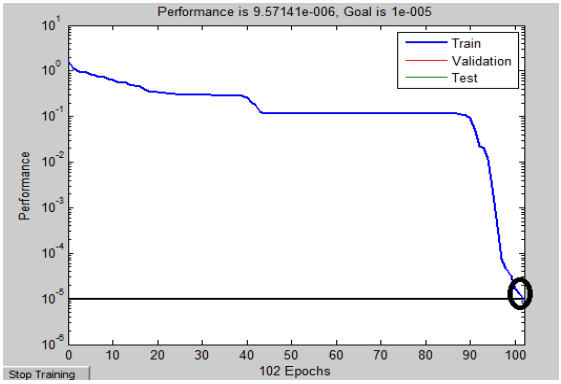

Fig 3: Performance curve (mean squared error) of training data set using Levenberg–Marquardt training algorithm

4.2 Training Result

In the proposed ANN based breast cancer tumor classification a total of 35 samples were used for training out of which include 15 benign and 23 malignant samples. The desired target value used for the classification was "0" for benign and "1" for malignant type tumor. Figure 4 shows the desired output and actual output during training.

4.3 Test and Validation results

Once a Neural network architecture is trained and set for number of inputs, number of hidden layer neurons, number of output layer neurons, transfer function for hidden layer and transfer function for output layer, the same Neural Network is used for testing. For the proposed system a total of 18 samples (10 samples corresponding to benign breast tumors and remaining 8 samples belonging to malignant Tumor) were used for testing performance of ANN based Breast Cancer Tumor Classification. . During testing phase it was found that there was one error i.e. sample number 5 is of benign category but the proposed system classified it as a malignant. Figure 5 shows the result during testing phase. The proposed system gives 94% accuracy.

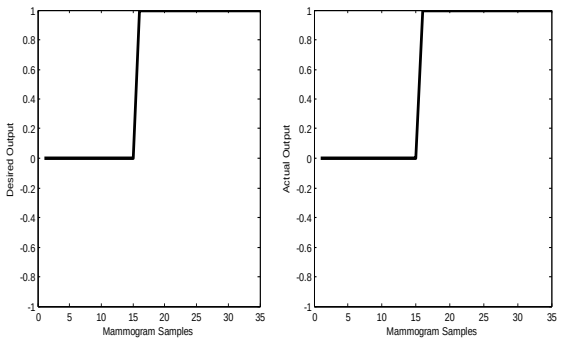

Fig 4: Set targets and actual outputs of ANN based Breast Cancer Tumor Classification during training

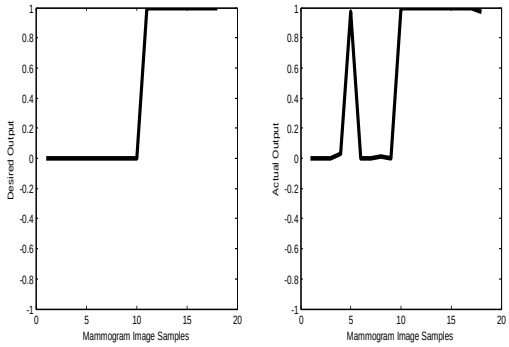

Fig 5: Test result of ANN based breast cancer tumor Classification

In validation phase 19 samples were taken for which the proposed system gives 94 percent accuracy. Figure 6 shows the result of proposed system during validation phase. In validation phase there was an error that was associated with sample number 7. The desired output for this sample was 0 but system gives one.

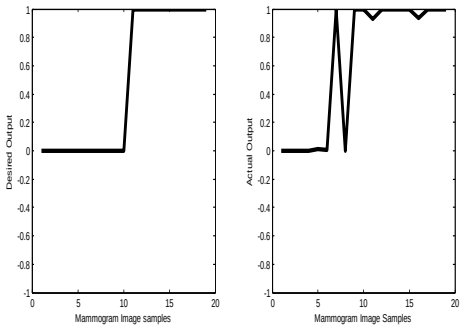

Fig 6: Validation result of ANN based breast cancer tumor Classification

5. CONCLUSIONS

In this paper an CBIR system and automatic method for classification of breast tumors in to malignant or benign using texture features is proposed. After extracting features, a one dimensional feature vector consisting of texture attributes was used for training and testing. ANN using back propagation algorithm was used for classification of the tumor. Results show that texture features can give satisfactory result in analysis and classification of mammographic images. ANN based breast cancer tumor classification gives 94 percent accuracy in classifying cancerous mammogram image as benign or malignant. The performance of the system can be further improved by fusing shape and texture features like GLRLM texture features etc. Further work is to check the performance of the system by increasing the size of database.

6. REFERENCES

[1] J. Ferlay, F. Bray, P. Pisani and D.M. Parkin. GLOBOCAN 2000: Cancer Incidence, Mortality and Prevalence Worldwide. Version 1.0. IARC Cancer Base No. 5. Lyon, IARC Press, 2001.

[2] Hala Al-Shamlan, Ali El-Zaart," Feature Extraction Values for Breast Cancer Mammography Images", proceedings of International conference on Bioinformatics and Biomedical Technology (ICBBT), pp. 335 – 340, 2010.

[3] B.K.Singh," Mammographic Image Enhancement, Classification and Retrieval Using Color, Statistical and Spectral Analysis, IJCA, 27(1):pp. 18-23, August 2011.

[4] J Suckling et al (1994): The Mammographic Image Analysis Society Digital Mammogram Database Exerpta Medica. International Congress Series 1069 pp375-378.

[5] Tabar L., Dean B.," Teaching Atlas of Mammography", 2nd edition, Thieme Newyork (1985).

[6] American Cancer Society, "Cancer Facts and Figures 2003", Atlanta, GA: American Cancer Society, 2003

[7] K.Rajakumar, Dr.S.Muttan ,"Medical Image Retrieval using Energy Efficient Wavelet Transform", Second International conference on Computing, Communication and Networking Technologies, Chennai,.India (2010).

[8] Y. Liu, D.S. Zhang, G. Lu and W.-Y. Ma, "Region-based image retrieval with perceptual colors, Proceedings of the Pacific-Rim Multimedia Conference (PCM)," December 2004, pp. 931–938.

[9] R. Shi, H. Feng, T.-S. Chua and C.-H. Lee, "An adaptive image content representation and segmentation approach to automatic image annotation," International Conference on Image and Video Retrieval (CIVR), 2004, pp. 545–554.

[10] V. Mezaris, I. Kompatsiaris and M.G. Strintzis, "ontology approach to object-based image retrieval," Proceedings of the ICIP, vol. II, 2003, pp. 511–514.

[11] Y. Ireaneus Anna Rejani, " Early Detection of Breast Cancer using SVM Classifier Technique", International Journal of Computer Science and Engineering Vol.1(3), pp. 127-130, 2009.

[12] B. K. Singh, G. R. Sinha, A. Wany and B. Mazumdar, Content based retrieval from biomedical images using color histograms, proceedings of International Conference on Nanotechnology and Biosensors, ICBN , Raghu College of Engg., Visakhapatnam , India 2010.

[13] R. M. Haralick, K. Shanmugam, and I. H. Dinstein, "Textural Features for Image Classification," IEEE Transactions on Systems, Man and Cybernetics, vol. 3, pp. 610-621.

[14] R. M. Haralick, "Statistical and structural approaches to texture," Proceedings of the IEEE, vol. 67, pp. 786-804, 1979.

[15] Scott Lindell, Gray Shapiro, Keneneth Weil, David Flannery, Jeffrey Levy, Wei Qian, " Development of Mammogram Computer – Aided Diagnosis Systems Using Optical Processing Technology", proceedings of workshop on Applied Imagery Pattern Recognition pp 173 - 179 , 2000.

[16] Tulio C.S.S. Andre and Rangaraj M. Rangayyan, " Classification of Tumors and Masses in Mammograms Using Neural Networks with Shape and Texture Features", Proceedings of 25th Annual International Conference on, pp. 2261 - 2264 Vol.3, 2003.