ISSN:0975-9646

Efficient CAD system based on GLCM & derived feature for diagnosing Breast Cancer

Rabi Narayan Panda[#1], Mirza Ashad Baig[*2], Dr. Bijay Ketan Panigrahi[#3], Dr. Manas Ranjan Patro[#4]

[#1] *Associate Professor, Krishna Institute of Engineering and Technology, Ghaziabad, UP* [#2]
Btech, Babu Banarasi Das National Institute of Technology and Management,Luckniw [#3]
Associate Professor, Indian Institute of Technology, Delhi, India.,
[#4] *Professor, Dept of Computer Science, Berhampur University,Berhampur, India*

Abstract : Breast cancer screening is done by performing mammography. According to statistics mammography diagnostic tests fails to detect up to 30% of breast lesions and up to 2/3 of those lesions are visible during reconsideration. With the advancement of medical technology, Computer Aided Diagnosis (CAD) has brought a revolutionary change in the areas of medical Imaging and Analysis and has expedited the early detection and diagnosis of cancerous tumours present in breast region. This paper proposes advanced Gray-Level-Co-Occurrence Matrix (GLCM) feature for textural feature from the segmented mammograms and efficiency is tested using three different classifiers Radial Basis Function Neural Network (RBFNN), Support Vector Machine (SVM) and KNN. A total of 112 images from MIAS database were trained and 74 images were used as test case, among all the three classifiers RBFNN with 93 percent gave the best efficiency. SVM and KNN followed with 88.73% and 86.4% percentages respectively. Confusion matrix and roc plot was also derived, by this way GLCM features were accurately able to detect the malignant lesions present in ROI . Thus, this paper instigates in increasing the probability of detection of cancerous breast tumor, avoiding delays in starting the treatment.

Keywords **— CAD, Gray-Level-Co-Occurrence Matrix (GLCM), ROI, RBFNN, SVM, KNN.**

INTRODUCTION

Cancer or tumor located in the breast region or tumor which originates in the breast is called as breast cancer. Defined into two groups based on the location, where first one is Ductal carcinoma and second one is Lobular carcinoma. The Ductal carcinoma initiates in ducts, ducts aid to move milk to nipple and the other is lobules which produce milk. As per statistics

25% of women's are affected by breast cancer [1] and every 1 out of 8 women are in the danger of being diagnosed with breast cancer [2]. Even male have the possibility of getting breast cancer but the probability of a male being affected is minimal. Mammography is the set preferred screening test for early detection of breast cancer. Around 50 to 90 thousand women in America are usually wrongly diagnosed with Breast Cancer annually [4] .Thus this has led to development of a better and an efficient digital algorithm resulting in precise and accurate digital mammograms thereby reducing the no. of cases occurring due to false positive results [3]. An extensive amount of skill is required to understand the complex problem encountered in breast cancer. So developing methods to reduce false positive cases is indispensable. [5]. Mammogram image generally has gray levels showing contrast which characterizes if it is a normal tissue or has calcification with masses. The tissue which appears white and opaque is generally a normal tissue and a fatty tissue has darker appearance. Figure 1(a) (b) & (c) shows normal, benign and malignant tissue respectively. Most of the time, breast lump is benign which means there is less danger of nearby cells also being affected.

Every image is represented by a texture which has a unique feature. So feature extraction, and its evaluation is a component which directly influences the output in mammogram classification. The most opted and result oriented features are Laws' texture energy [6], spatial gray-level dependence [7], Fourier power spectrum [8]. Researchers have dedicated lot of time in finding the best feature and in improvising the classifier's efficiency to classify tumors detected in mammograms as benign and malignant. Various CAD systems have been developed and are used in obtaining a second opinion. Due to the complexity in identifying the texture in some cases an efficient algorithm for estimating the cancer is required Rangayyan et al [9-11] proposed the method based on Gabor filters and phase portrait maps to characterize oriented texture patterns in mammograms. Brake et al [12] presented method based pixel orientation. A technique based on Linear filter for enhancement of speculation was proposed by sampat et al [13]. A method based on asymmetry was proposed by Giger et al.[14]. A method on textural feature based on multichannel filtering was proposed by Ole et al [15]. According to the breast type using law texture mask Bovis and Singh [16] proposed a method where texture energy was retrieved for the use of feature. As pointed out by Zwiggelaar[17] favorable outcomes in various application is often provided by GLCM, Thus in this paper , we have evaluated the performance of 3 Different classifiers, RBFNN , SVM, KNN (with textural feature achieved by GLCM) for the classification of tissue,

(a) : Normal (b) : Benign (c) : Malignant

Figure 1 : Mammogram images

with the objective of having accuracy in identification of type of tumor and segregation of malignant from benign tissues, GLCM is a quite commonly used statistical method for extracting textural feature from various digital images.

II. METHODOLOGY

As breasts have a piece wise structure texture, therefore feature extraction signifies a distinguishable and suitable feature to classify the selected ROI. The features generally have reference to neighbourhood operation or some specific structure in image. Based on the spatial variation of the neighbouring pixels with certain mathematical derivations GLCM features are extracted. Conventionally used mammogram images are highly textured and complex which makes interpretation very difficult .Harlick et al [18] had proposed a method of GLCM features for textural feature extraction , Due to the complexity of the data used an approach based on second order statistical method was formulated. It is based on the probability of finding a gray level pair at random distance with different orientation over an image ROI. GLCM features are generated from the intensities of pairs of pixels where spatial relationship is detailed in the form of distance and angle . A matrix is generated based on number of pixel pairs having grey level values. The steps used in feature extraction to classification is as discussed below

2.1 *Preprocessing*

For extracting features and training the data as benign and malignant, roi must be defined, the marker is provided in the database as x , y co-ordinate which forms the centre and r defines the radius .for further implementation this has to be converted to a rectangular roi which is defined by equation (1).

$$I_{ROI}=I\,[x-r,\ 1000-y,\ 2r,\ 2r] \qquad \text{EQ 1}$$

The image can then be resized for uniformity into 128 x 128, 512 x512, and 256 x 256. In this work the ROI were resized to 64 x 64.

2.2 *Feature extraction*

We have followed certain notation like p(i,j) is (i,j)th entry for a normalized gray tone matrix . where px(i) is the ith entry for the probability matrix which has been obtained from summing rows of p(i,j) which is given by equation (2)

$$P(I,J),= \sum_{j}^{Ng} = 1^{P(i,j)} \qquad \text{EQ2}$$

Here Ng is total number of distinct gray levels in the image and mean value is represented by μ for P. Similarly for Px and Py means and standard deviation are given by . μx, μy, σx, σy.. The features are as mentioned below:

$$E = \sum_{i,j} P(i,j)^{2} \qquad \text{EQ.3}$$

$$P=\text{MAX}_{I,J}\, P(I,J) \qquad \text{EQ.4}$$

$f_4 = \sum_i \sum_j (\mu)(1-\mu)^2 p(i,j)$. EQ5

$f_7 = \sum_{i=2}^{2N_g} (i - f_8)^2 \; p_{(x+y)}(I)$. EQ .6

f_{10} =VARIANCE OF $p_{(x-y)}$ EQ.7

we also propose some new matrices from the GLCM matrix which have shown efficient result for classifying benign and malignant . We have denoted the new matrices by F_{new} and F_{new1} given in the equation 11 and 12

$F_{NEW} = \sum_{i,j} P(i,j)^2$ + MAX $_{I,J}$ $P(I,J)$ + VARIANCE OF $p_{(x-y)}$ + $\sum_{i=2}^{2N_g} i p^2_{(x+y)^2}$ …EQ11

A threshold is set by the optimization algorithm which shows less the is the value of combined index(f_{new}) of ROI, chances of detection of malignancy is lower . F_{new1} is another metric defined by square root of sum of variance to difference of variance

F_{NEW1} = SQRT($\sum_i \sum_j (\mu)(1-\mu)^2 p(i,j)$) / VARIANCE OF $p_{(x-y)}$) ….EQ.12

The extracted features are normalized and the optimal features are trained to the classifier.

*2.3 K-Nearest Neighbor (KNN) Classifier***:**

The basic crux is to use the majority rule. that assigned point to the class is used as a sample point from which the majority of the k nearest neighbors belong. when classifying to more than two groups or when using an even value for k, it might be necessary for the tie in the number of nearest neighbors. selection of random options results in selection of random tie breaker, and 'nearest', which uses the nearest neighbor among the tied groups to break the tie. the default behavior is majority rule, nearest tie-break. the distance measured are usually euclidean, mahalanobis, cosine, correlation,spearman,hamming,jaccard or a custom distance function.

2.4 Support Vector Machine (SVM) Classifier:

A Support Vector Machine (SVM) is a classifier which formally defined by a distinguishing a hyperplane. The labeled training data (*supervised learning*), the defined algorithm results into an optimal hyperplane which classifies new examples. The distinctive advantage of an SVM is that it results in a unique result . SVMs are that that have a understandable geometric explanation and produces a crisp solution. The computational complexity of SVMs does not depend on the size or dimensions of space in which the input is provided. Structural risks are drastically reduced with the help of SVMs. This is the prime reason why SVM as a classifier works better than conventionally designed ANNs.

2.5 Radial Basis Function Neural Network (RBFNN) :

In order to discover the potential image of micro calcification, mass lesions in the breast tissue images a proper reliable method must be used.[19]In our proposed CAD technique we used RBFNN as the potential classifier. In RBFNN the value of the real valued function depends only on the origin distance. That is, if a function 'h' satisfies the property h(x)=mod(h(x) then it is known as a radial function. Its characteristic feature response increases or decreases monotonically with center point distance. Number of hidden neurons selected is 54 for RBFNN to generate a better efficiency for the dataset.

RESULTS

This work focused on developing an efficient CAD system using KNN, SVM and RBFNN classifiers. This work utilized 112 images; out of which 56 were malignant. The segmentation of mammogram for removing pectorial muscle and x-ray annotation is presented in our previous work [20] . Seven GLCM features were found efficient to classify mammogram images into benign and malignant. KNN classifier gave an efficiency rounded of 86% with svm being 88% and RBFNN gave an efficiency of 93% . Features used for training is given in section 2.2 It was analyzed as the number of features increases, the performance with the classifier would be better but the computational complexity would also increase. Performance evaluation of the classifier can be done by analyzing sensitivity, specificity and positive predictive accuracy (PPA). The sensitivity of a test is the probability that it will produce a true positive result (A) when used on an affected subset of population. To evaluste the performance of the classifier ROC curve has been shown for all classifier. Fig 2 shows ROC for KNN with fig 3 for SVM and fig 5for RBFNN

GLCM	TP	TN
TP	21	7
TN	3	43

Table 1: Confusion Matrix for KNN classifier

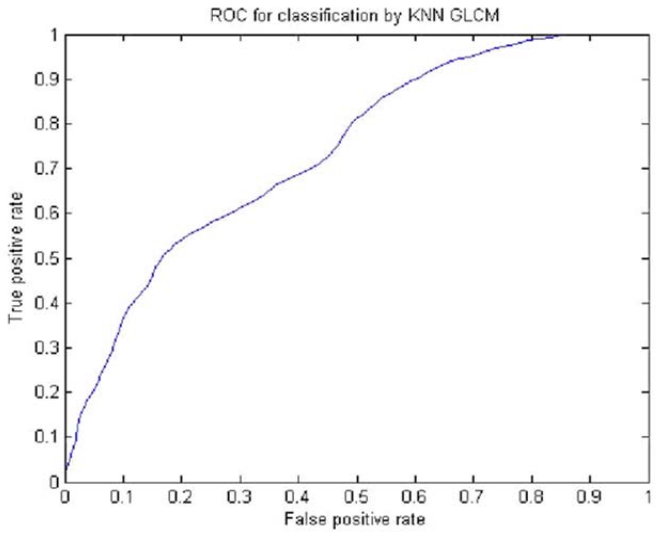

Figure 2 : ROC curve for knn classifier

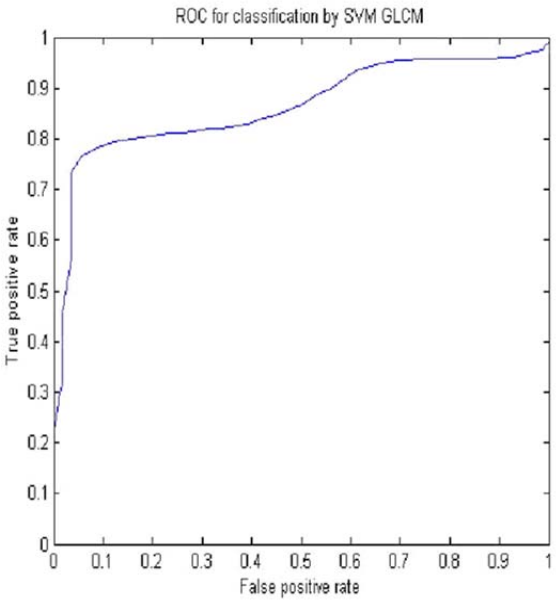

Figure 4: ROC curve for svm

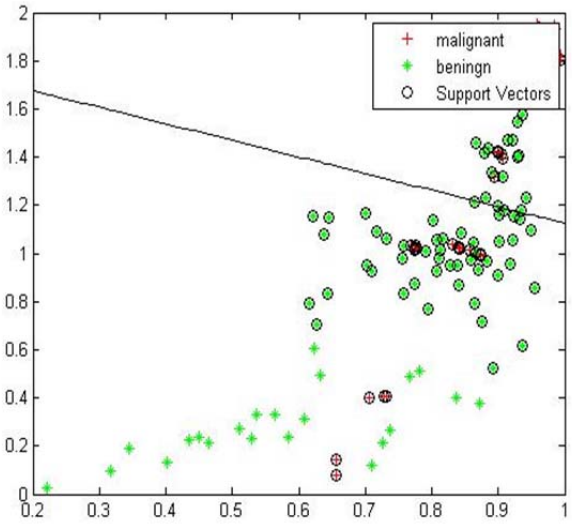

Figure 3: Decision boundary for SVM Classifier

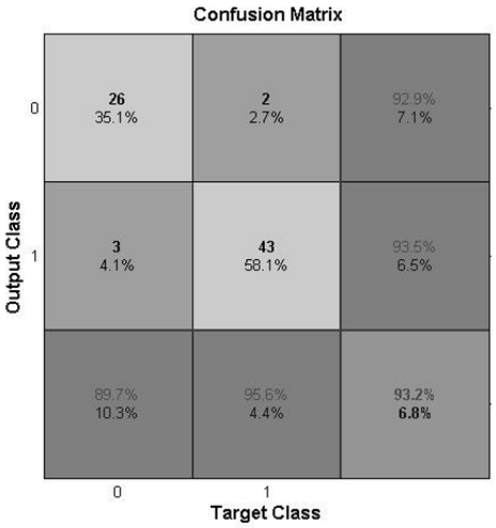

Figure 5 : Confusion matrix for RBFNN

GLCM	TP	TN
TP	21	6
TN	2	42

Table 2 : Confusion Matrix for SVM classifier

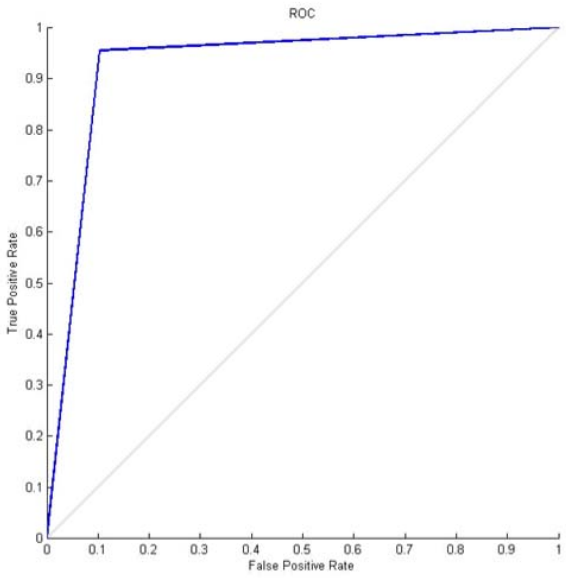

Figure 6 : ROC CURVE FOR RBFNN

CONCLUSION

This work presented a CAD system which can help radiologists in identifying whether the solid breast nodule is malignant or benign. The objective of this work was to compare the performance of three different classifiers RBFNN,KNN,SVM with the textural features derived from GLCM and also validate the efficiency of new texture defined by us in the work. RBFNN is appreciable with good accuracy for GLCM textural features . A very important step for the RBFNN training is to decide the proper number of hidden neurons. If the number of hidden neurons does not chosen properly, the RBFNN may show poor global generalization characteristics, slow training speed and the need for large memory requirement. Therefore the correct number of RBF neurons and appropriate cluster distance factor should be considered carefully while designing the RBFNN for classification. The simulation results show strong evidence of effectiveness in early detection of breast cancer in mammograms.

REFERENCES

[1] Chalasani P, Downey L, Stopeck AT. Caring for the breast cancer survivor: a guide for primary care physicians. Am J Med. 2010;123(6):489-95.

[2]. Cuzick J, DeCensi A, Arun B, et al. Preventive therapy for breast cancer: a consensus statement. Lancet Oncol. 2011;12(5):496-503 .

[3] Maggie Mahar," Health Beat –Commentry on Health Care,the Economy,politics,Public Health

[4] M. Elter, R. Schulz-Wendtland and T. Wittenberg .The prediction of breast cancer biopsy outcomes using two CAD approaches that both emphasize an intelligible decision process. Medical Physics , American Association of Physicists in Medicine , 2007. 34(11): 4164-4172.

[5] Gordon D. Schiff, Seijeoung Kim, Richard Abrams, Karen Cosby, Bruce Lambert, Arthur S. Elstein, Scott Hasler, Nela Krosnjar, Richard Odwazny, Mary F. Wisniewski, Robert A. McNutt,"Diagnosing Diagnosis Errors: Lessons from a Multi-institutional Collaborative Project", Vol 2 Advances in Patient Safety- Diagnosing Diagnosis Errors

[6] Daniel B, Md Kopans, "Breast Imaging", Lippincot Williams, 1997.

[7] Pradeep N, Girisha H, Sreepathi B, Karibasappa.k, " Feature extraction of mammogram" IJBR , Vol. 4, Issue 1, 2012, pp.-241-244.

[8] C. S. Burrus, R. A. Gopinath, and H. Guo, Introduction to Wavelets and Wavelet Transforms: A Primer. Upper Saddle River, New Jersey: Prentice Hall, 1998.

[9] Oliver. Automatic mass segmentation in mammographic images. PhD Thesis. Department of Electronics, Computer Science and Automatic Control. University of Girona,2007.

[10] L. Giger, F. Yin, and K. Doi. Investigation of methods for the computerised detection and analysis of mammographic masses. SPIE Medical Imaging and Image Processing IV,1233:183 }184, 1990.

[11] P. Miller and S. Astley. Classification of breast tissue by texture analysis. Image and Vision Computing, 10:277{283, 1992.

[12] J.Juhl. Paul and juhl Essentials of Roentgen Interpretation. 4th edition Harper, pages 340-345, 1982.

[13] L. Blot, R. Zwiggelaar, and C.R.M. Boggis. Enhancement of abnormal structures in mamographic images. Proceedings of Medical Image Understanding and Analysis, pages 125-128, 2000.

[14] L. Giger, F. Yin, and K. Doi. Investigation of methods for the computerised detection and analysis of mammographic masses. SPIE Medical Imaging and Image Processing IV,1233:183{184, 1990.

[15] T. Ole, Gulsrud, and E. Loland. Multichannel filtering for texture extraction in digital mammograms. 18th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 1996.

[16] K. Bovis and S. Singh. Detection of masses in mammograms using texture features. 15th International Conference on Pattern Recognition (ICPR'00), 2000.

[17] L. Blot and R. Zwiggelaar. Extracting background texture in mammographic images: a co-occurrence matrices based approach. Proceedings of the 5th International Workshop on Digital Mammography, Toronto(Canada), pages 142-148, 2000.

[18] . R.M.Haralick, K.Shanmugan, and I.H.Dinstein, Texture Features for image classification," IEEE Trans.Syst., Man, Cyber., vol. SMC-3, pp.610-621, 1973.

[20] Rabi Narayan Panda, Mirza Ashad Baig, Dr. Bijay Ketan Panigrahi,, Dr. Manas Ranjan Patro,' An Efficient X-ray annotation removal Algorithm and cluster based segmentation for mammogram Images', International Journal of Scientific & Engineering Research, Volume 6, Issue 4, April-2015