AUTOMATED DETECTION OF TUMOURS IN MAMMOGRAM – A COMPARISON BETWEEN RBF AND SVM

C. Sathish Kumar[#1], R. Arun Kumar[*2], K. Babu[*3], D. Dinesh Kumar[*4], M.S. Jaffer Sathik[*5]
[#]*Assistant Professor, Department of Biomedical Engineering,*
Adhiyamaan College of Engineering, Hosur-635109, Tamilnadu.
[1] csathishkumar08@gmail.com
[*]*Student, Department of Biomedical Engineering,*
Adhiyamaan College of Engineering, Hosur-635109, Tamilnadu
[2]arun220593@gmail.com

Abstract-**Diagnostic Mammograms are used in the diagnosis of breast disease in women with breast symptoms and abnormal screening results. Screening mammograms usually take 2 views of each breast; while diagnostic mammograms may take more views of the breast. Calcifications which are tiny mineral deposits can be micro calcifications and macro calcifications. The proposed work compares the performance of Support Vector Machines and Radial Basis Function and their accuracy in detection of lesions in mammographic images and elimination of false positives. Tumor in mammograms is identified using morphological operation and the abnormality is classified using GLCM features and RBF and Support Vector Machine supervised neural network classifier. The images from the data set are initially pre-processed and contrast enhanced which makes the image effective for further analysis. Then Region Of Interest (ROI) is determined using threshold based tumour segmentation by Otsu method. Various features like Gray Level Co-occurrence Matrix (GLCM) features, Discrete Wavelet Transform (DWT) features are derived for the particular ROI. Radial Basis Function (RBF) classifier and Support Vector Machine are trained with the features using MATLAB bioinformatics tool box. Thus the classified results are obtained for the input image based on the trained RBF structure. The mammography data set has been downloaded from the Mammographic Image Analysis Society (MIAS) in which the images are available along with ground tooth information.**

Key Words- **Mammography, Gray Level Co-occurrence Matrix (GLCM), Discrete Wavelet Transform (DWT), Support Vector Machine (SVM), Radial Basis Function (RBF).**

I. INTRODUCTION

With much advancement in medical field, the reason for breast cancer is still unknown. Breast cancer is the leading cause of cancer deaths among women all over the world and according to researches mortality in case of breast cancer accounted for around 13.7% of all cancer deaths in women in 2008. Considering the incidence rates for women in India in the same year the rate was estimated to be 22.9%. The incidence to mortality ratio was around 2:1. It is significantly higher than that of the United States for the same period. Regular screening of breasts in women above the age of 40 can significantly reduce mortality rates. Mammography is considered the gold standard and the most reliable method of detection till date.

Therefore mammograms play a vital role in early diagnosis of breast cancer and helps in controlling it and taking the necessary therapeutic efforts. The nature of tissues in mammography images derives the main complexity in identifying the type of cancer. Usually, the X-ray component of a mammogram is required for breast cancer screening purposes. A lesion usually appears brighter than the surrounding normal tissue on a mammogram. This is because the area denser than fat stops more x-rays photons. Most of the researchers have chosen digital mammography for their magnification of an area, brightness of the film may be adjusted after the examination is completed, enabling the radiologist to see certain areas more clearly.

Cancers vary in their appearance and size depending upon the progression of the disease. Computer aided diagnosis is an important tool in assisting doctors. CAD provides another dimension to doctors' point of view, thereby reducing the chances of missing out positive cancerous lesions. . Hence it is unavoidable to go for some automatic CAD approaches that do not involve manual intelligence.

It is very difficult for the physicians and radiologists to analyze between a malignant and benign mass. Recent studies show that still there is about 1-5% of misclassification and false positives. This study is based on the previous research classification approaches that resulted in good accuracy rates in classifying benign and malignant tumors. The results obtained were analyzed for its efficiency using some of the performance metrics like sensitivity, accuracy and precision value in the case of neural networks.

Calcifications are tiny mineral deposits within the breast tissue that appear as small white spots on the films. They may or may not be caused by cancer. There are 2 types of calcifications:

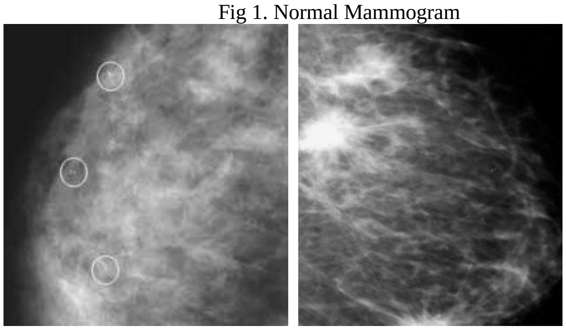
Macrocalcifications are coarse (larger) calcium deposits that most likely represent degenerative changes in the breasts, such as aging of the breast arteries, old injuries, or inflammation. These deposits are associated with benign (non-cancerous) conditions and do not require a biopsy. About half the women over the age of 50, and in about 1 in 10 women younger than 50, have macrocalcifications.

Microcalcifications are tiny specks of calcium in the breast. They may appear alone or in clusters. Microcalcifications seen on a mammogram are of more concern than macrocalcifications, but do not always mean that cancer is present. The shape and layout of microcalcifications help the radiologist judge how likely it is that cancer is present. In most instances, the presence of microcalcifications does not mean a biopsy is needed. If the microcalcifications look suspicious for cancer, a biopsy will be done.

Lesions that occupy space can be classified into masses, architectural distortion (ARCH) and asymmetry (ASYM). The masses can be further classified into circumscribed masses (CIRC), speculated masses (SPIC) and other masses depending upon the shape and marginal features. Space occupying lesions are often indistinguishable from the surrounding mass which is usually glandular tissue because of similarity in attenuation properties.

Lesions with smooth margins are usually benign while malignant masses show speculated boundaries that develop gradually over time. Usually masses range from 3mm to 50mm and at times they appear blurred. Therefore we apply initial preprocessing steps in order to prepare the image for further processing.

Fig 1. Normal Mammogram



2 (a) 2 (b)

Fig 2. (a) Mammogram with Microcalcifications
(b) Mammogram with Macrocalcifications

II. Review of previous Methods

Mudigonda N. R. et al proposed a method for breast mass detection in mammograms which was initially done by density slicing and texture flow field analysis and this method uses Gaussian smoothing and sub sampling operations as preprocessing steps[7]. The mass portions are segmented by establishing intensity links from the central portions. Features based on flow orientation in adaptive ribbons of pixels across the margins of masses are proposed to classify the regions detected as true mass regions or false-positives (FPs). The mass regions that were successfully segmented were further classified as benign or malignant disease by computing five texture features based on gray-level co-occurrence matrices (GCMs) and using the features in a logistic regression method. The result of this method was determined to have an accuracy of 0.87 with mass versus normal tissue classification under receiver operator characteristics.

Mavroforakis M. E. et al have done characterization based on localized texture analysis of breast tissue on mammograms based on dataset fractal analysis using linear, neural and support vector machine classifiers. However, in contrast to other mammographic diagnostic approaches, it has not been investigated in depth, due to its inherent difficulty and fuzziness [3]. Establishment of a quantitative approach of mammographic masses texture classification is based on advanced classifier. Textural features are extracted at larger scales and sampling box sizes prove to be more content-rich than their equivalents at smaller scales and sizes. Fractal analysis on the dimensionality of the textural datasets verified that reduced subsets of optimal feature combinations can describe the original feature space adequately for classification purposes and at least the same detail and quality as the list of qualitative texture descriptions provided by a human expert.

Aize Cavo et al employed robust information clustering incorporating spatial information for breast mass detection in digitized mammograms [1]. It employs RIC algorithm based on the raw region of interest (ROI) extracted from global mammogram by two steps of adaptive thresholding. Pixels on the fuzzy margin of a mass and noisy data were identified by RIC through the minimax optimization of mutual information. The memberships of the identified pixels (outliers) were recalculated by incorporating spatial distance information that takes into account of the influence of a neighborhood of 3×3 window. Only suspicious regions are located and further examination by experts is required and elimination of false positives by experts is essential.

Abhijit Nayak et al have proposed a method for detection of suspicious lesions using undecimated wavelet transform and adaptive thresholding. This is used to generate a multiresolution representation of the original mammogram. Adaptive global and local thresholding techniques are then applied to segment possible malignancies. False positive elimination by experts is needed to identify lesions in initial stages [8].

Automatic detection of tumour subtype in mammograms based on GLCM and DWT feature using SVM was done by M.Mohamed Fathima et al [4]. It states that this method increases the accuracy of classification and reduces the percentage of false positives. The images from the data set are initially preprocessed and contrast enhanced which makes the image most effective for further analysis. Then Region Of Interest (ROI) is determined from morphological top hat filtered image by means of thresholding segmentation. Various features like first order textural features, Gray Level Co-occurrence Matrix (GLCM) features, Discrete Wavelet Transform (DWT) features, run length features and higher order gradient features are derived for the particular ROI. Support Vector Machine (SVM) classifier is trained with the above mentioned features using MATLAB bioinformatics tool box. Thus the classified results are obtained for the query image based on the trained SVM structure. The statistical risk of misclassification must be minimized by maximizing the margin between the support vectors and the hyper plane. Enhanced can be done by extracting shape features and by using clustering algorithms for segmenting the ROI. Overlapping of features using SVM classifier must be avoided.

III.METHODOLOGY

The proposed method uses Local Otsu Thresholding method for effective segmentation and compares the performance of Radial Basis Function Neural Network and Support Vector Machine based on spectral features. The series of steps involved in this method include thresholding using Otsu method and morphological processing on the image and detection of tumour while at the same time we extract feature from the input image and this feature is given to the trained neural network. The neural network is trained using the database and once the trained neural network classifies the image the result is obtained.

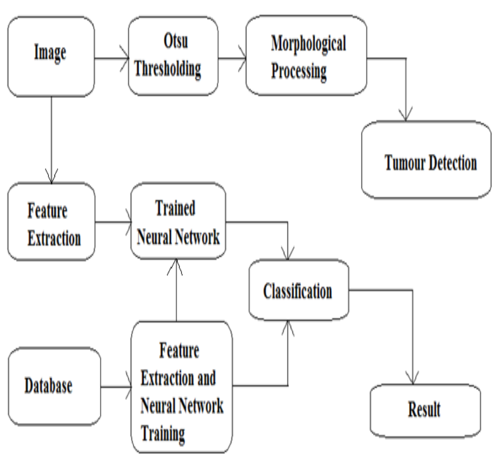


Fig 3. Flow Diagram

1. Preprocessing and Segmentation :

The image is pre processed in order to reduce the probability of error and increase the speed of processing. The steps involved include initial resizing, cropping and scaling and then enhancing the contrast of the image. Enhancement is the process of manipulating an image in order to make it more suitable for further processing.

The contrast of the image is improved using the formula:

$$f_{cont}(x, y)= \frac{f^2(x,y)*255}{\max\{f^2(x,y)\}}$$

Equation: 1

Mammogram is enhanced using top hat filtering algorithm and with a disk shaped structuring element of size 12. The filtered image is processed with local thresholding algorithm for detecting the region of interest [5]. One obvious way to extract the objects from background is to select a threshold T, then at any point (x, y) in the image at which f(x, y)>T is called an object point otherwise it is a background point. Thus, the segmented image g(x, y) is given by

$$g(x, y)=\begin{cases} 1 \;\; if\; f(x,y) > T \\ 0 \;\; if\; f(x,y) \le T \end{cases}$$

Equation: 2

when T varies throughout the image depending upon the properties of the neighbourhood pixels the method is denoted as Local thresholding. Due to its intuitive properties, simplicity of implementation and computational speed, image thresholding enjoys a central position in applications of image segmentation. Otsu's method is known to be optimum as it maximizes the between-class variance, which is used in statistical discriminant analysis. This method is based entirely on computations performed on the histogram of the image.

2. Feature Extraction:

The input data must be transformed into the set of features is done in order to classify the data. The choice of features that are extracted must be chosen carefully since it must perform the desired classification using only the selected features instead of the entire ROI. Large amount of memory and computational power is required provided there is ample amount of information. The features extracted include textural features, higher order gradient features and using Daubechies, Haar filters discrete wavelet transform features are also extracted. Textural feature extraction involves two steps which include determination of co-occurrence matrix and then calculating the features based on it. A Co-occurrence matrix calculates specific spatial relationship between two pixels. The spatial relationship is defined as the pixel of interest. The number of gray levels in the image determines the size of the co-occurrence matrix.

3. Training of Support Vector Machine and Radial Basis Function:

Support Vector Machine which is a supervised classifier that analyses data for classification with associated learning takes a set of input data and predicts the possible output. The hyperplane acts as the decision boundary of the classifier. SVM is effective in these cases where the number of dimensions is greater than the number of samples. The features extracted using GLCM method and DWT are applied to Radial Basis Function network. This network which consists of three layers which includes the input, hidden and output layer uses Gaussian activation function. Once the network is trained with extracted features the network is ready to classify the images. The performance of this network can be evaluated using specificity, sensitivity and accuracy.

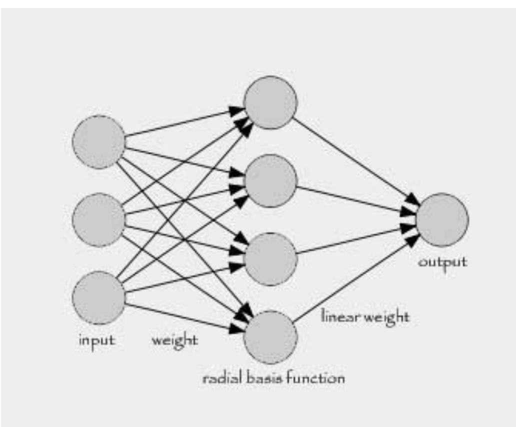


Fig 4. Radial Basis Function Network

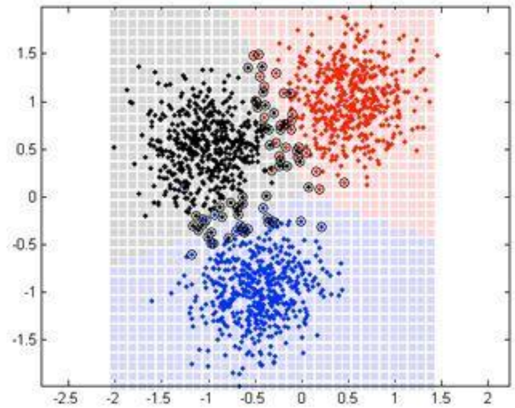


Fig 5. Multiclass Support Vector Machine Classifier

IV.RESULTS AND DISCUSSION

Till date mammography is considered to be the best diagnostic modality which is least invasive and helpful in detecting breast cancer. Detection of tumours might be tedious in most of the cases. In order to make the classification of result simple we go in for these automated detection methods which involve least human intervention. By making the segmentation results automatic we result in better accuracy. In order to improve segmentation results we have improvised local thresholding instead of global thresholding. Support Vector Machines are used in order to obtain acceptable results and one inevitable criterion that results in this process is the problem of overlapping. Better result possibility is known to exist when shape features are extracted in even more accurate methods using clustering algorithms. Using Radial Basis Function networks also yield results that give better accuracy than feed forward neural networks.

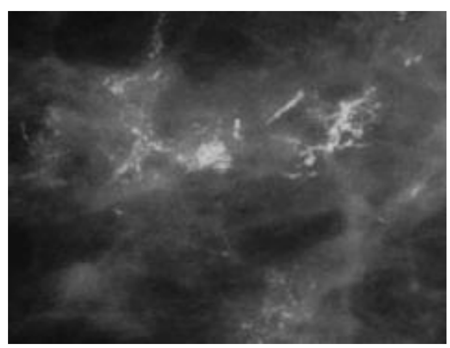


Fig 6. Result produced by RBF classifier

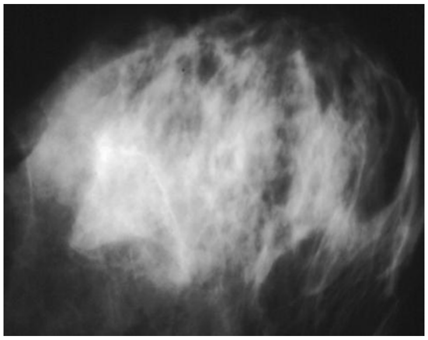


Fig 7. Result produced by SVM classifier

V. CONCLUSION

According to supervised classification techniques that we have compared Radial Basis Function yields accurate results depending upon the features extracted and improvement is essential as far as accuracy and sensitivity is considered. While Support Vector Machines can classify the tumours far better depending upon the classes for which they have been trained. Improvements in future on this work include improvisingnesting characteristics into support vector machines so that they have the ability to reduce overlapping of data and produce fine detailing upon the images to be classified and obtain greater accuracy and sensitivity in results.

REFERENCES

[1] Aize Cao, Qing Song and Xulei Yang, "Robust information clustering incorporating spatial information for breast mass detection in digitized mammograms," Computer Vision and Image Understanding, Vol. 109, No. 1, pp. 86-96, 2008.

[2] Matthew A. Kupinski and Maryellen L. Giger, "Automated Seeded Lesion Segmentation on Digital Mammograms," IEEE Transactions on Medical Imaging, Vol. 17, No. 4, pp. 510-517, 1998.

[3] Mavroforakis ME, Georgiou HV, Dimitropoulos N, Cavouras D and Theodoridis S, "Mammographic masses characterization based on localized texture and dataset fractal analysis using linear, neural and support vector machine classifiers,"

[4] M.Mohamed Fathima, Dr.D.Manimegalai, and Ms.S.Thaiyalnayaki, "Automatic Detection of Tumor Subtype in Mammograms Based On GLCM and DWT Features Using SVM," ICICES, 2013.

[5] P. Mohanaiah and Dr. P. Sathyanarayana, "Detection Of Tumour Using Grey Level Co-Occurrence Matrix And Lifting Based DWT With Radial Basis Function," IJERT, Vol. 2, No. 6, pp. 1677-1684, 2013.

[6] Moi Hoon Yap, Eran A. Edirisinghe, and Helmut E. Bez, "A novel algorithm for initial lesion detection in ultrasound breast images," Journal of Applied Clinical Medical Physics, Vol. 9, No. 4, 2008.

[7] Mudigonda N.R. and Rangayyan R. M, "Detection of breast masses in mammograms by density slicing and textural flow field analysis," IEEE Transactions on Medical Imaging, Vol. 20, No. 12, 2002.

[8] Nayak, Abhijit, Ghosh, Dipak Kumar, Ari and Samit, "Suspicious lesion detection in mammograms using undecimated wavelet transform and adaptive thresholding," ICACT, 2013.

