Detection of Breast Cancer in Mammogram using Support Vector Machine

Varsha J. Gaikwad

Savitribai Phule University, Pune, Maharashtra, India

Abstract: *Breast cancer is leading cause of death in women. An early sign of breast cancer is the presence of a tiny deposit of calcium i.e microcalcifications in mammogram. To detect this micro-calcification the Support Vector Machine (SVM) based classifier techniques is used. In this paper we proposed the technique to detect the cancer in the mammogram. The proposed method has been implemented in four stages preprocessing, segmentation for ROI extraction, feature extraction and classification. The proposed method was evaluated with Mammogram Image Analysis Society (MIAS) database. Experimental results show that, SVM shows better accuracy.*

Keywords: Breast Cancer, Mammography, preprocessing, marker-controlled segmentation, SVM classifier

1. Introduction

Breast cancer is caused by the development of malignant cells in the breast. The malignant cells originate in the lining of the milk glands or ducts of the breast (ductal epithelium), defining this malignancy as a cancer. Cancer cells are characterized by uncontrolled division leading to abnormal growth & the ability of these cells to invade normal tissue locally or to spread throughout the body, in a process called metastasis.

Mammography is the diagnostic screening procedure to detect breast cancer in the female breast using low-dose X-rays.The difference in absorption of X-rays between the various tissue components of the breast such as fat, tumor tissue, and calcifications. If mammography is not sufficient, other techniques can be used such as ultrasonography and MRI. In the process of screening, the mammographic image is called mammograms are made from different angles (views or projections). The most common projections of the breast are the mediolateral oblique (MLO) view & the craniocaudal (CC) view.Breast abnormalities are defined with wide range of features & may be easily missed or misinterpreted by radiologists while reading large amount of mammographic images provided in screening programs. There is a difficulty to detect masses in mammograms because sometimes masses seemed to be similar to normal breast tissues on mammograms. It is difficult to distinguish between malignantand benign masses. Irregular shapes have a higher probability of being malignant and regular shapes have a probability of being benign. Difference in regions of the right and left breast is known as bilateral asymmetry of the breast [1]. Thus, we have used the combination of different techniques on mammogram to detect the breast cancer which will be helpful to radiologists.

R. Ramani et all [2] has been research on the preprocessing techniques for breast cancer detection in mammography images. They were research on median , adaptive median , mean & wiener types of filtering are used for pre-processing to improve image quality, remove the noise, preserves the edges within an image, enhance and smoothen the image, mainly concentrate the MSE, PSNR and AE. Finally, compared the simulated output parameters such as image quality, mean square error, Peak signal to noise ratio, structural content and normalized absolute error on 322 mammogram images (MIAS).

D. Sujitha Priya et all [3] research on breast cancer detection in mammogram images using region-growing and contour-based segmentation techniques by the implementation of pre-processing methods such as, mean filtering, median filtering and adaptive median filtering.Adaptive Median Filtering technique that is implemented with a Median filter produced the best result among three with measuring MSE and PSNR value.

Jawad Nagi et al [4] have developed an algorithm on artifact suppression & background separation.Raw mammogram image contains wedges and labels. These may produce unnecessary disturbances during mass detection process. Hence it should be removed in preprocessing. They were proposed the method thresholding and morphological opening, closing, dilation and erosion are used to remove these artifacts.

Armen Sahakyan [5] has developed an algorithm on Segmentation of the Breast Region in Digital Mammograms & Detection of Masses **.**In mammogram images radiopaque artifacts such as wedges & labels are removed using threshold technique and morphological operations for enhancement purpose.

R. Subash Chandra Boss et all [6] research on automatic mammogram image breast region extraction and removal of pectoral muscle. The presence of pectoral muscle in mammograms may disturb thus he has proposed on automated method to identify the pectoral muscle in MLO view mammograms based on histogram based 8-neighborhood connected component labelling method for breast region extraction and removal of pectoral muscle. The proposed method is evaluated by using the mean values of accuracy and error. The comparative analysis shows that the proposed method identifies the breast region more accurately.

2. Breast Cancer Detection

2.1 Preprocessing

In our proposed method the first step is involves preprocessing usinf Median filter. In this, the image is enhanced and smoothed.It is a nonlinear filter which efficient in removing salt and pepper noise median tends to keep the sharpness of image edges while removing noise [7].
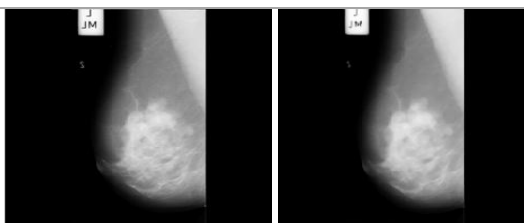
Figure 2.1: a: Mdb181 1024 pixels x 1024 pixels after Resizing into 256 x256, b: Median Filter

2.2 Artifact Suppression & Pectoral Muscles Removal

In this research the MIAS image dataset has been used.Raw mammogram image contains wedges & labels. These may produce unnecessary disturbances during mass detection process. Hence it should be removed in preprocessing. In the proposed method thresholding & morphological opening, closing, dilation & erosion are used to remove these artifacts.

Radiopaque artifacts such as wedges & labels in the mammograms images are removed using thresholding & morphological operations. From the result, larger area should be extracted neglecting the small areas. Due to the above process smaller artifact regions can be easily removed from the mammogram. Morphological erosion &dilation is applied using structuring elements. Finally the holes produced in the binary image are filled using imfill function in Matlab. Resulting image is artifacts separated image.

Similar to the label and artifact, the mammogram image contain the pectoral muscle which is also produce unnecessary disturbances during mass detection process. The pectoral is the term relating to the chest. It is a largest muscle that covers much of the front upper chest. The pectoral region should always appear as a high intensity, triangular region across the upper posterior margin of the image. In many cases the upper part of the boundary is a sharp intensity edge while the lower part is more likely to be a texture edge, due to the fact that it is overlapped by fibro glandular tissue [8]. Hence during the mammogram capturing process pectoral muscle also would be captured it will severely affect the result of image processing. Hence it is also removed before the segmentation stage in ROI detection.

Radiopaque artifacts such as wedges, labels and breast profile separation in the mammograms images are removed using Thresholding and morphological operations [9] .The proper selection of Thresholding value should be important otherwise it will threshold maximum portion of image. Thus to overcome this problem a connected component labeling method is being used.

2.2.1 8-Neighborhood Connectivity:

In the 8-neighborhood connectivity in that pixels are connected if their edges or corners touch. This means that if two adjoining pixels are on, they are part of the same object, regardless of whether they are connected along the horizontal, vertical, or diagonal direction.
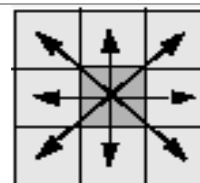

Figure 2.2: 8-neighborhood connectivity

Figure 2.3: a) Binary image of Pectoral Muscles b) Pectoral Muscles Removal

2.3 Marker-controlled Watershed Segmentation

The watershed transform is a tool morphological based for image segmentation. Separating touching objects in an image is one of the more difficult image processing operations. So the watershed transform is often applied to this problem. Watershed represents the boundaries between adjacent catchments [10]. The minimum can be interpreted as markers of watershed regions.
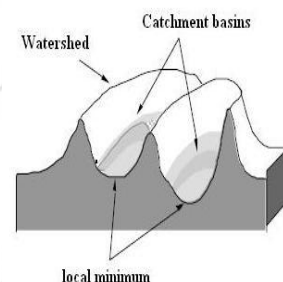
Figure 2.4: Watershed lines with catchments basins

The basic principle of watershed technique is to transform the gradient of a grey level image in a topographic surface, where the values of f(x, y) are interpreted as heights and each local minimum embedded in an image is referred as catchments basins.
By using watershed transform 'catchment basins' and 'watershed ridge lines' in an image by treating it as surface where light pixels are high and dark pixels are low.

Figure 2.5: Final Segmented Image

2.4 Feature Extraction

Feature extraction is a method of capturing visual content of a segmented image. The objective of feature extraction process is to represent raw image in its reduced form to facilitate de-

International Journal of Scientific Engineering and Research (IJSER)
www.ijser.in
ISSN (Online): 2347-3878
Volume 3 Issue 2, February 2015

cision making process such as pattern classification [11]. A Statistical intensity based features are extracted from the mammograms for classification. Feature extraction step is important step to get high classification rate. A set of features are extracted in order to allow a classifier to distinguish between normal & abnormal pattern. The abnormality can be identified on the basis of textural appearance. Extracted features are used in Support vector machine classifier to train it for the recognition of particular class either normal or abnormal. The ability of the classifier to assign the unknown object to the correct class is dependent on the extracted features [12].

Let, L be the number of possible intensities in an image $M \times N$ pixels, z_i, i = 0:,1, 2,…L − 1 their intensity values, and n_i the absolute frequence that z_i occurs in the image. P is the probability of Zi occurs in the image.

1. Mean:
It is the average value of intensity of the Image.It is defined as,

(1) $$\mu = \sum_{i=0}^{L-1} z_i P(z_i)$$

2. Standard Deviation:
It is the square root of the variance.The standard deviation is the estimate of the mean μ square deviation of gray pixel value P (Zi) from its mean value.The Standard deviation is defined as,

(2) $$SD = \sqrt{\sigma^2} = \sqrt{\sum_{i=0}^{L-1}(Zi - \mu^2)P(Zi)}$$

3. Entropy:
Entropy is a statistical measure of randomness that can be used to characterize the texture of the input image.It is defined as,

(3) $$E = -\sum_{i=0}^{L-1} P(Zi)\log_2[P(Zi)]$$

4. Skewness:
Skewness, S characterizes the degree of asymmetry of a pixel distribution in the specified window around its mean. It is of a distribution is defined as,

(4) $$s = \frac{E(Z-\mu)^3}{\sigma^3}$$

5. Kurtosis:
Kurtosis, K measures the Peakness or flatness of a distribution relative to a normal distribution.It is defined as,

(5) $$k = \sum_{i=0}^{L-1}(Zi - \mu^4)P(Zi)$$

6. Variance:
The average of squared differences from the mean. Variance is the square root of standard deviation. It is defined as,

(6) $$\sigma^2 = \sum_{i=0}^{L-1}(Zi - \mu^2)P(Zi)$$

7. Energy:
Energy is a measure of the uniformity.
It is defined as sum of square of pixel values of element.

It is defined as:

(7) $$E = \sum_{i=0}^{L-1}[P(Zi)]^2$$

8. Correlation:
The operation called correlation is closely related to convolution. In correlation, the value of an output pixel is also computed as a weighted sum of neighboring pixels.
It is is defined as:

(8) $$C = \frac{\sum_i \sum_j P(i,j) - \mu_x \mu_y)}{\sigma_x \sigma_y}$$

9. Smoothness:
Relative smoothness, R is a measure of grey level contrast that can be used to establish descriptors of relative smoothness. The smoothness is defined as:

(9) $$S = R = 1 - \frac{1}{1+\sigma^2}$$

Where, σ is standard deviation.

10. RMS: Root Mean Square:
The RMS (Root Mean Square) computes the RMS value of each row or column of the input, along vectors of a specified dimension of the input, or of the entire input. The RMS value of the jth column of an M-by-N input matrix u is given by below equation:

(10) $$y = \sqrt{\frac{\sum_{i=1}^{M}\left|u_{ij}\right|^2}{M}}$$

2.5 Support Vector Machine

Support vector machines (SVMs) are a set of related supervised learning methods that analyze data and recognize patterns, used for classification and regression analysis. The standard SVM takes a set of input data, and predicts, for each given input, which of two possible classes the input is a member of, which makes the SVM a non-probabilistic binary linear classifier [13]. Since an SVM is a classifier, then given a set of training examples, each marked as belonging to one of two categories, an SVM training algorithm builds a model that predicts whether a new example falls into one category or the other. More formally, a support vector machine constructs a hyperplane or set of hyperplanes in a high or infinite dimensional space, which can be used for classification, regression or other tasks. Intuitively, a good separation is achieved by the hyperplane that has the largest distance to the nearest training data points of any class (so-called functional margin), since in general the larger the margin the lower the generalization error of the classifier.

The basic idea of SVM is that it projects data points from a given two-class training set in a higher dimensional space and finds an optimal hyperplane. The optimal one is the one that separates the data with the maximal margin. SVMs identify the data points near the optimal separating hyperplane which are called support vectors. The distance between the separating hyperplane and the nearest of the positive and negative data points is called the margin of the SVM classifier.

The separating hyperplane is defined as

(11) $D(x)=(w.x)+b$

Where x is a vector of the dataset mapped to a high dimensional space, and w and b are parameters of the hyperplane that the SVM will estimate. The nearest data points to the maximum margin hyperplane lie on the planes
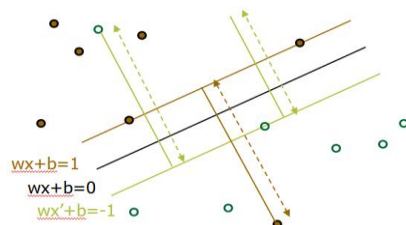

Figure 2.6: Representation of Hyper planes

Where y=+1 for class w_1 and y=-1 for class W_2.

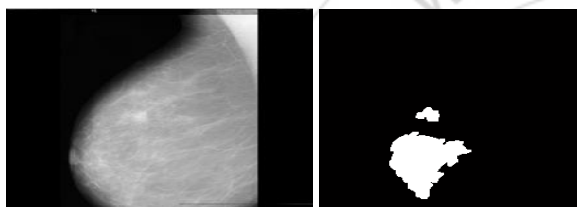The width of the margin is given by $m = \dfrac{2}{\|w\|}$.

Figure 2.7: a) Original Image b) Segmented Image c) Status of Mdb265

3. Database

Total numbers of 322 samples are obtained from mammograms containing microcalcifications from Mammographic Image Analysis Society (MIAS). These were divided into training and testing set. The training set contains 285 samples, of which, 255 normal, benign and malignant cases. Remaining 30 samples are used for testing SVM classifier. Testing set contains 15 cases of normal and 15 cases of abnormal respectively.

3 Receiver Operator Characteristic

The Receiver Operating Characteristics (ROC) curve of the implemented classifier can be plotted by providing several values of sensitivity and specificity based on SVM classification having accuracy 83%.
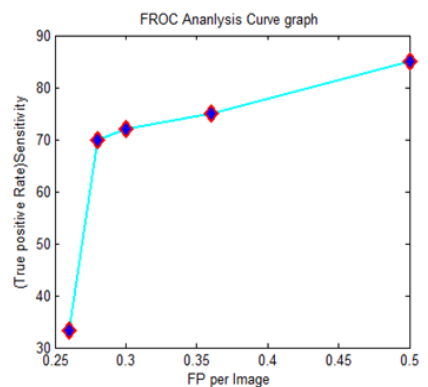

Figure 2.8: FROC Curve

4. Conclusions

Mammography is one of the best methods in breast cancer detection, but in some cases, radiologists cannot detect tumors despite their experience. Such computer-aided methods like presented in this paper could assist medical staff and improve the accuracy of detection.

In this paper, we present a classification of mammograms using SVM classifier. The proposed classification scheme has been applied on MIAS database image. Experimental results show that, high accuracy 83% of classification of MCCs obtained with the proposed SVM classifier can help radiologists in making an accurate diagnostic decision, which can reduce unnecessary biopsies. Our future work is to extend the various different features and other classification type of mammogram.

5. Acknowledgments

Every work is an outcome of full-proof planning, continuous hard work and organized effort. This work is a combination of all the three put together sincerely.

References

[1] Jawad Nagi , Sameem Abdul Kareem , Farrukh Nagi , Syed Khaleel Ahmed "Automated Breast Profile Segmentation for ROI Detection Using Digital Mammograms", 2010 IEEE EMBS Conference on Biomedical Engineering & Sciences (IECBES 2010), Kuala Lumpur, Malaysia, pp87-99,2010.

[2] R. Ramani "The Pre-Processing Techniques for Breast Cancer Detection in Mammography Images" I.J. Image, Graphics and Signal Processing, pp 47-54, 2013.

[3] D. Sujitha Priya "Breast Cancer Detection In Mammogram Images Using Region-Growing And Contour-Based Segmentation Techniques" International Journal of Computer & Organization Trends ,Volume 3, Issue 8 Sep 2013,ISSN: 2249.

[4] Jawad Nagi,1, Sameem"Automated Breast Profile Segmentation for ROI Detection Using Digital Mammograms"2010 IEEE EMBS Conference on Biomedical Engineering & Sciences (IECBES 2010), Kuala Lumpur, Malaysia, 30th November - 2nd December 2010.

[5] Armen Sahakyan "Segmentation of the Breast Region in Digital Mammograms and Detection of Masses" Interna-

tional Journal of Advanced Computer Science and Applications, Vol. 3, No.2, 2012.

[6] R. Subash Chandra Boss, K. Thangavel, D. Arul Pon Daniel " Automatic Mammogram image Breast Region Extraction and Removal of Pectoral Muscle".

[7] Pragathi. J, H. T. Patil "Segmentation Method for ROI Detection in Mammographic Images using Wiener Filter and Kittler's Method" International Journal of Computer Applications (0975 – 8887) International Conference on Recent Trends in engineering & Technology , pp27-33, 2013.

[8] S. Shanthi, V. Murali Bhaskaran "Computer Aided System for Detection and Classification of Breast Cancer", International Journal of Information Technology, Control and Automation (IJITCA) Vol.2, No.4, October 2012.

[9] Jaime S. Cardoso, Ines Domingues, Amaral, "Pectoral muscle detection in mammograms based on polar coordinates and the shortest path", 32nd Annual International Conference of the IEEE EMBSBuenos Aires, Argentina, August 31 - September 4, 2010.

[10] K. Parvati, b. S. Prakasa rao, and m. Mariya das"Image Segmentation Using Gray-Scale Morphology and Marker-Controlled Watershed Transformation", hindawi publishing corporation Discrete dynamics in nature and society volume 2008, article id 384346, 8 pages-doi:10.1155/2008/384346,2008.

[11] Zaheeruddin , Z. A. Jaffery a, Laxman Singh "Detection and Shape Feature Extraction of Breast Tumor in Mammograms" Proceedings of the World Congress on Engineering 2012 Vol II ,WCE 2012, July pp4 - 6, London, U.K.,2012.

[12] R. Ramani "The Pre-Processing Techniques for Breast Cancer Detection in Mammography Images" I.J. Image, Graphics and Signal Processing, pp 47-54, 2013.

[13] Zhiqiong Wang, GeYu, YanKang, YingjieZhao, QixunQu "Breast tumor detectionin digitalmammography basedon extreme learning machine", Neurocomputing 128(2014)175–184, 2013 Elsevier