Breast Cancer Detection Using SVM and Minimum Distance Classifier

Dr. S.R Ganorkar[1], Dipali A. Sable[2]
[1, 2] Dept of Electronics and telecommunication
[1, 2] SCOE, Pune

Abstract- *Breast cancer is leading cause of cancer deaths among women in India. It starts to the breast and can spread to the other parts of the body. Screening mammography is the most reliable method available for early detection of breast cancer. The disease is curable if detected early. Research indicates that the mortality rate could decrease by 30% if women age 50 and older have regular mammograms. Digital mammograms have become the most effective techniques for the detection of breast cancer. The goal of this research is to increase the diagnostic accuracy of image processing and machine learning techniques for optimum classification between malignant and benign abnormalities in digital mammograms by reducing the number of misclassified cancers.*

In this proposed work, the mammogram images are initially preprocessed using different methods. In this, noise in the background will be removed using median filter and wiener filter and contrast enhancement will be done using contrast limited adaptive histogram equalization techniques.

Then the region of interest will be determined using segmentation by otsu's thresholding algorithm. Features of the mammogram images will be extracted using Wavelet Transform and the extracted features are used for the classification. Support Vector Machine and normalize minimum distance classifier are used to classify the images.

Keywords- Mammogram images, minimum distance classifier, otsu's thresholding, Support vector machine, wavelet transform

I. INTRODUCTION

The Breast Cancer is leading cause of death among females. It is a second reason of death of women after lung cancer [7].

Medical imaging techniques are widely used in the diagnosis of breast cancer to monitor progression. False diagnosis can be reduced by using computers due to increase in accuracy.

Mammography is the imaging technique used to examine breasts uses a low dose X-ray system. But there is a possibility of indentifying the abnormality that looks like a cancer but it turn out to be normal which will be called as a false negative report by using mammography. Such misdiagnosis will lead to more tests which would be more stressful to the patient.

The crossing number method was implemented before, for finding the cancer cell area but that method worked only in the image preprocessing steps. But for final conformation the doctors should go for biopsies or other tests to get conclusion on their diagnosis process which give much more pain to the patients. So there is a need to work more on image post processing step to find cancer area from the image itself which is very effective for diagnosing the cancer cell.

A. Objective

The objective of this proposed work is to identify and diagnose breast cancer at its early stage. The following issues are addressed-

- To study the breast cancer analysis using image processing.
- To remove the noise and image enhancement using preprocessing.
- To perform segmentation and feature extraction with accuracy.
- Developing a framework of algorithms using image processing and machine learning techniques for the detection of malignant and benign abnormalities in digital mammography images.
- To classify normal and abnormal images accurately using machine learning algorithm.

II. LITERATURE REVIEW

Sangeetha R. and Dr. Srikanta Murthy K, 2017 their paper, "A novel approa**olqch for detection of breast cancer at an early stage using Digital Image Processing techniques", presents the novel techniques for the detection of breast cancer using preprocessing, segmentation using Otsu's thresholding

algorithm, feature extraction using gray level co-occurrence matrix and classification using Baye's classifier[1].

Jisna Jose, Ms. Anusha Chacko, and D.Anto Sahaya Dhas,2017, "comparative study of different image denoising filters for mammogram preprocessing", they explains and compared different types of filters[2].

Rajesh Garg, Bhawana Mittal, Sheetal Garg 2011, "Histogram Equalization techniques for Image Enhancement explains different methods of histogram equalization", with flowcharts and compared it using different parameters[3].

B.K.Gayathri, P.Rajan, 2016, "A Survey of Breast Cancer detection based on image segmentation techniques", explains different types segmentation methods for mammography images. There methods and compared according to its accuracy, speed,etc[4].

Abdul Quyyam, A.Basit,2016, "Automation Breast Segmentation and cancer detection via SVM in Mammograms", gives the information about segmentation and detection of microcalcification cells using support vector machine classifier[5].

Nagla S. Ali Ibrahim, Nagla F. Soliman,Mahmoud Abdullah, and Fathi E. Abd El-Samie,2016,their paper "An algorithm for pre-processing and segmentation of mammogram images", they focused on preprocessing and segmentation of the mammogram imges. Preprocessing is mainly using mdian filter, local thresholding and histogram equalization methods. Segmentation is done using Otsu's thresholding algorithm[6].

M.M.Fathima, Dr. D. Manimegalai, Ms.S.Thaiyalnayaki, "Automatic detection of tumor subtype in mammograms based on GLCM and DWT Features using SVM", explains feature extraction methods that is gray level co-occurrence matrix and discrete wavelet transform[7].

Otsu N., "A Threshold Selection Method from Gray-level Histograms," explains the Otsu's thresholding algorithm for segmentation.[8]

Issam El-Naqa, Yongyi Yang, Miles N. Wernick, Nikolas P. Galatsanos,2002,"A Support Vector Machine Approach for Detection of Microcalcifications" investigate an approach based on support vector machines (SVMs) for detection of microcalcification(MC) clusters in digital mammograms, and propose a successive enhancement learning scheme for improved performance[9].

Robin N. Strickland, Hee I1 Hahn,1996,"Wavelet Transforms for Detecting Microcalcifications in Mammograms", explains the basics of wavelet transform and gives the subband explaination. It explains the multiresolution criteria and biorthogonal wavelet transform[10].

Issam El-Naqa, Yongyi Yang,2002, "A Support Vector Machine Approach for Detection of Microcalcifications", in this the proposed SVM framework outperformed all the other methods tested. In particular, a sensitivity as high as 94% was achieved by the SVM method at an error rate of one false-positive cluster per image. The ability of SVM to outperform several well-known methods developed for the widely studied problem of MC detection suggests that SVM is a promising technique for object detection in a medical imaging application[11].

Ernest L. Hall, Richard P. Kruger,1971, "A Survey of Preprocessing and Feature Extraction Techniques for Radiographic Images" explains Several preprocessing techniques for enhancing selected features and removing irrelevant data are described and compared. The techniques include gray level distribution linearization, digital spatial filtering, contrast enhancement, and image subtraction. Also, several feature extraction techniques are illustrated. The techniques are divided into spatial and Fourier domain operations. The spatial domain operations of directional signatures and contour tracing are first described[12]

III. DATASET COLLECTION

The mammography case samples required for the study has been taken from the Mammographic Image Analysis Society (MIAS). The Mammographic Image Analysis Society (MIAS) is an organization of UK research groups interested in the understanding of mammograms and has generated a database of digital mammograms. It contains 322 mammogram images of size 1024 × 1024 pixels with ground truth information about the abnormalities, i.e., type of cancer, severity of the diagnosis (Benign or Malignant), centre coordinates of location of the abnormality and radius of the circle enclosing the abnormality.

IV. METHODOLOGY

The proposed methodology consists of following techniques

1. Image Preprocessing
2. Image segmentation
3. Feature extraction
4. Feature classification

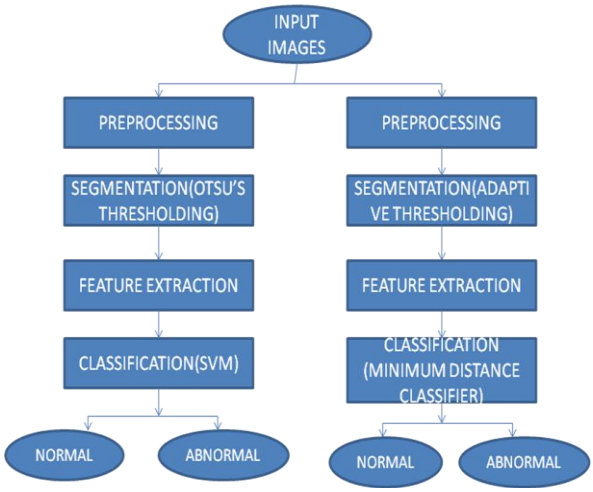


Figure1. Flow diagram of the proposed system

1. Image Preprocessing

Image preprocessing is important to remove the noise and to enhance the quality of image. In this stage regions of interest are enhanced and the unwanted regions of the image are deemphasized. The enhancement of the regions of interest and the suppression of noise is performed in this stage. This can be achieved by using Median filter in first method and weiner filter in second method .

Median filter is used to remove the salt and pepper noise and weiner filter is useful to remove speckle noise. Contrast limited adaptive histogram equalization algorithm is used increase the contrast of the image.

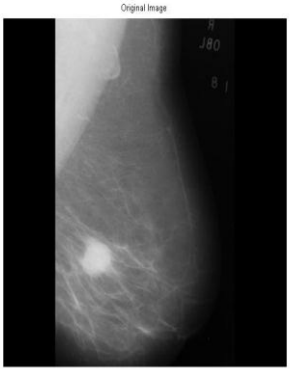


Figure2(a): Original image

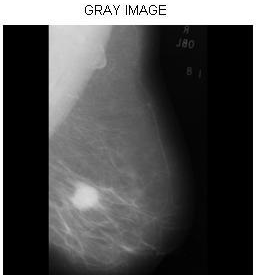
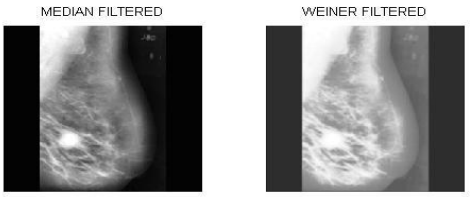


Figure2(b): Grey image



2. *Segmentation*

Image segmentation is the process of subdividing the image into its constituent parts to obtain the required object from the background. Segmentation is done based on discontinuity and similarity of pixels. Discontinuity includes identification of isolated points or edges or lines. Similarity includes grouping of similar pixels.

In this work segmentation is done using otsu's thresholding.

2.1Otsu's thresholding

a. Compute the histogram of the image. Let each gray level have probability pi.

b. Calculate the aggregate sum for k = 0,1,…….L-1

$$P_1(K) = \sum_{i=0}^{k} P_i$$

c. Calculate the aggregate mean $m_1(k)$ for k = 0 to L-1

$$m_1(k) = \frac{1}{P_1(k)} \sum_{i=0}^{k} ip_i$$

d. Calculate the overall intensity mean

$$m_G = \sum_{i=0}^{L-1} ip_i$$

e. Calculate variance between the class $\sigma_B^2(k)$ for k = 0,.....,L

$$\sigma_B^2(k) = P_1(k)(m_1(k) - m_G)^2 + P_2(k)(m_2(k) - m_G)^2$$

f. Find the value of k for which $\sigma_B^2(k)$ is maximum.

Otsu's thresholding algorithm will segment the microcalcification from the image with a high accuracy.

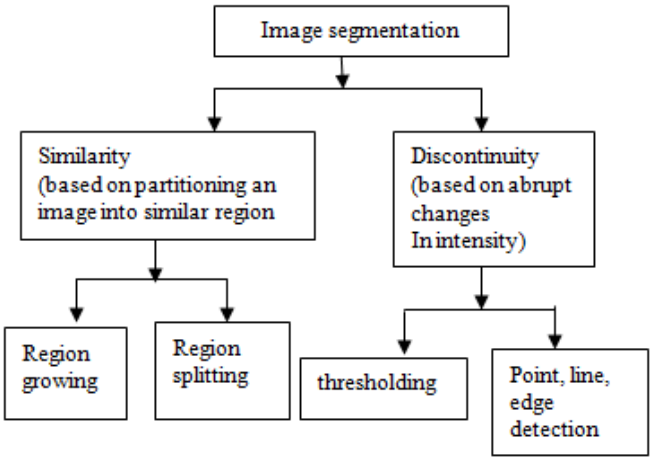


Figure3(a): Image segmentation classification

Figure3(b): ROI obtained after segmentation

3. Feature Extraction

Features of the segmented region are extracted in this stage such as size, shape and texture. The extracted feature must be carefully chosen because desired classification task is expected to perform using this representation instead of using complete region. In this work features are extracted to classify type of breast tissue to distinguish normal and cancerous breasts using haar wavelet transform and biorthogonal wavelet transform. These extracted features is used for the classification.

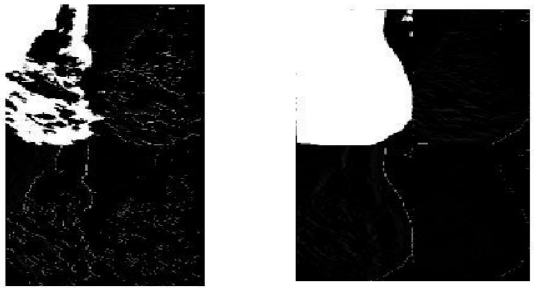


Figure 4: Wavelet feature using haar wavelet and biorthogonal wavelet

TABLE 1: WAVELET FEATURES

wavelet Features	Values
Contrast	7.7646
Correlation	0.8948
Energy	0.118
Homogeneity	0.208

4. Feature Classification

Classifier is a mathematical model used to classify the regions of interest into different classes. The extracted features are used to classify the mammogram images into normal and abnormal with high accuracy. Support Vector Machine and minimum distance classifier is used to in this work.

SVM is the supervised machine learning classification technique that is widely used in the cancer diagnosis. SVM functions by selecting critical samples from all classes which is known as support vectors and separated by generating linear function.

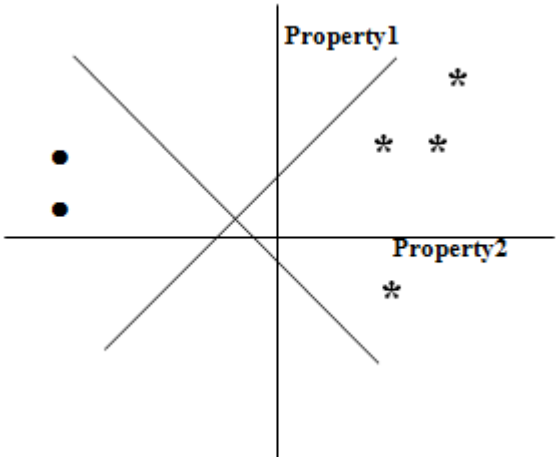


Figure 5: SVM generated hyperplanes

The minimum distance classifier is used to classify unknown image data to classes which minimize the distance between the image data and the class in multi-feature

space. The distance is defined as an index of similarity so that the minimum distance is identical to the maximum similarity.

$$d_{x,m} = \sqrt{x^2 - m^2}$$

x is the tested pixel.
m is the mean value of the cluster.

V. PERFORMANCE CRITERIA

In Mammography image classification techniques, the performance metrics to be determined are Accuracy, Sensitivity and Specificity. The sensitivity of a test is the proportion of people with have the disease who test positive for it. The specificity of a test is defined as the proportion of Victims without the disease who will test negative for it.

1. True positive: Victims correctly diagnosed as sick (TP)
2. False positive: Fit people incorrectly diagnosed as sick (FP)
3. True negative: Fit people correctly diagnosed as healthy (TN)
4. False negative: Victims incorrectly diagnosed as healthy. (FN)

Accuracy = (TP+TN)/(TP+TN+FP+FN)
Precision=TP/(TP+FP)
Sensitivity=TP/(TP+FN)

VI. CONCLUSION

TABLE 2: CONFUSION MATRIX 1

CLASSIFIED GROUPS	ACTUAL GROUPS		TPR	TNR
	ABNORMAL	NORMAL		
ABNORMAL	16(TP)	4(FN)	80	
NORMAL	2(FP)	78(TN)		97.5

TABLE 3: CONFUSION MATRIX 2

CLASSIFIED GROUPS	ACTUAL GROUPS		TPR	TNR
	ABNORMAL	NORMAL		
ABNORMAL	15(TP)	5(FN)	75	
NORMAL	30(FP)	50(TN)		62.5

Preprocessing is done using median and the wiener filter in which noise is removed. Otsu's thresholding is used for the segmentation. Features are extracted using wavelet transform. Results are observed using two methods. Total 100 images are used for the training in which 80 are negative and 20 are positive images.

- Using SVM:
  - From the confusion matrix it is observed that 16 positive images and 78 negative images are correctly diagnosed.
  - accuracy = 94%
  - TPR = 80%
  - TNR = 97.5%
- Using Minimum distance classifier:
  - In this method 15 positive images and 50 negative images are correctly diagnosed.
  - Accuracy = 65%
  - TPR =75%
  - TNR = 62.5%

Comparing both the methods accuracy using support vector machine is more than minimum distance classifier. Thus proposed methodology is completely automatic and helps to diagnose the mammogram images with accuracy and precision

REFERENCES

[1] Sangeetha R.,Dr. Srikanta Murthy K, "A novel approach for detection of breast cancer at An early stage using Digital Image Processing techniques", *International Conference on inventive systems and control*,(2017)
[2] Jisna Jose, Ms. Anusha Chacko, D.Anto Sahaya Dhas, "comparative study of different image denoising filters for mammogram preprocessing", *International conference on inventive systems and control*,(2017).
[3] Rajesh Garg, Bhawana Mittal, Sheetal Garg, "Histogram Equalization Techniques For Image Enhancement", *IJECT Vol.2, Issue 1, March 2011*.
[4] B.K.Gayathri, P. Raajan, "A Survey of Breast Cancer detection based on image segmentation techniques",2016.
[5] Abdul Qayyum, A. Basit, "Automation breast segmentation and cancer detection via SVM in mammogram", 2016.
[6] Nagla S. Ali Ibrahim, Nagla F. Soliman,Mahmoud Abdullah, Fathi E. Abd El-Samie, "An algorithm for pre-processing and segmentation of mammogram images",2016.
[7] M.M.Fathima, Dr. D. Manimegalai, Ms.S.Thaiyalnayaki, "Automatic detection of tumor subtype in mammograms based on GLCM and DWT Features using SVM".

[8] Otsu N., "A Threshold Selection Method from Gray-level Histograms," IEEE Transaction on Systems, Man & Cybernatic 1979,Volume 9, Issue 1 pp: 62-66.

[9] Robin N. Strickland, Hee I1 Hahn,"Wavelet Transforms for Detecting Microcalcifications in Mammograms*", IEEE Transactions on medical imaging, vol. 15, no. 2, April 1996*

[10] Issam El-Naqa, Yongyi Yang, "A Support Vector Machine Approach for Detection of Microcalcifications*", IEEE Transactions on medical imaging, vol. 21, no. 12, December 2002*

[11] Ernest L. Hall, Richard P. Kruger, "A Survey of Preprocessing and Feature Extraction Techniques for Radiographic Images", *IEEE transactions on computers, vol. C-20, no. 9, September 1971*

[12] Rafael C. Gonzalez, Richard E. Woods, "Digital Image Processing Using MATLAB".

[13] Rafael C. Gonzalez, Richard E. Woods, "Digital Image Processing".