Detection of Breast Cancer from Thermography Images

Suganthi L, Nirmala K, Shaalu Sree B.N, Pavithra S, Yashvantha N

Abstract: Breast cancer is now the most common cancer in most cities in India, and the second most common in rural areas. Early detection of breast cancer by systematic evaluation of the individual may improve survival rate. Infrared thermography one of the imaging technique that produce high resolution infrared images shows the heat pattern based on the temperature changes in breast with respect to the progression of the cancer cells. Increased metabolic activity and the blood flow due to the multiplication of cancer cells induces more heat on the skin layer which are captured by the thermal camera to produce the thermal images. This paper discusses on the real time image processing algorithm to detect the presence of cancer from the acquired thermal images. The methodology includes the preprocessing the acquired image and segmenting the region of interest, extracting the features from the segmented image followed by feature selection and classification. From the results, it is inferred that ANN classifiers yields better classification accuracy of 92% and minimum error rate (0.08) in K-means segmentation method when compared with SVM, KNN classifiers.

Index Terms: ANN classifier, CLAHE, IR Thermography, K-means clustering, KNN classifier, Otsu thresholding, SVM.

I. INTRODUCTION

As indicated by the Union health ministry, breast cancer positions as the main disease among Indian women with a rate of 25.8 per 100,000 women and mortality of 12.7 per 1,00,000 women. Thus, early detection and fast diagnosis of breast cancer is the key to save the patient's life and decrease mortality rate. At present the most common methods for detecting the breast diseases are Computed Tomography Laser mammography (CTLM), Mammography, Magnetic Resonance Imaging (MRI), Doppler Ultrasoundography, Positron Emission Mammography (PEM) etc. With the innovation in science and technology many medical imaging methods such as Ultrasound, MRI, PET, CT and few others were discovered. Medical imaging techniques are classified into two broad categories invasive, ionizing and non-invasive, non-ionizing. Due to the recent advancement in the thermal infrared technology, IR systems have received lots of attention for the use of the medical imaging especially in breast cancer detection. Therefore, a high concern has been given to thermography as an effective breast cancer screening tool.

Breast thermography is the non-invasive, non-radiating, quick and eases imaging approach for early discovery of breast disease. It is a successful approach for imaging for women independent of ages, sizes, and type of breast. The technique is additionally pertinent to screen the breast condition after medical procedure.

Breast thermography has some important advantages over mammography such as its ability to deal with dense breast tissues. Finally, the ionization, high pressure, and compression of breast are not required in thermography so no rupture risk here. Breast thermography can detect the indication of cancer earlier than mammography. Because of all the above reasons, digital thermal imaging is accepted as the most appropriate modality for earliest screening.

Infrared breast thermography is not a structural imaging technique which measures distribution of temperature that can be investigated. Many different image processing approaches have been used to automatically detect breast cancer in the literature.

Assessment of breast thermal image primarily depends on the texture features for detection of differences between the two breasts. Statistical parameters can be used for the asymmetry analysis of breast thermograms. Early research study shows the usefulness of various statistical parameters in the detection of breast cancer [17]. Effectiveness of texture information is evaluated by analyzing the extracted features in discriminating the abnormal thermal image from the normal thermal image [1]. One of the studies has used Otsu thresholding method to segment the region of interest. Initially the breast region extracted from the image followed by extraction of features. Extracted features are fed to classifier for performance analysis like accuracy, specificity and sensitivity [16].

Earlier research works discuss about the automated segmentation for breast cancer analysis using the optimal features and classification of normal and abnormal breast thermograms using optimal feature selectors and classifiers.

Research Studies on breast thermogram images show that Contrast Limited Adaptive Histogram Equalization (CLAHE) technique can be used for enhancing the preprocessed breast thermogram image and then it
will be segmented to extract the region of interest using k
means and fuzzy C means. Various features are extracted
from the segmented images. Finally a comparison has been
made by using the SVM and Bayesian classifiers [2], [3] and
[5]. The main objective of this work is to develop an
automated breast cancer detection algorithm, which
indicates the presence or the absence of the tumor in the
breast thermograms. Image processing techniques should
effectively extract important information from the suspicious
region and should classify the thermograms into tumor and
non-tumor cases. The image pre-processing and
segmentation algorithms are the essential requirements for
the relevant features extraction process from the breast
thermograms. Our objective is to develop an effective
analysis of breast cancer is determined and discussed. The
experimental results focus on the performance of the
classifier with high accuracy value. The results of those
methods, statistical, textural and shape based features have
been extracted from the segmented ROI image. Then,
significant features are identified and the performance
measures of the classifiers such as Artificial Neural Network
(ANN), Support Vector Machine (SVM) and KNN are used
to classify the normal and abnormal based on the selected
features.

II. METHODOLOGY
This section deals with the methodology, fig.1, that
constitutes of step by step procedure starting with the
pre-processing, then image segmentation followed by feature
extraction and feature selection, finally the optimized
features are used to detect the presence of breast tumor
present in the thermography images with help of strong
classifier.

A. Visual data base Acquisition
DMR-IR Database used for this work contains 72 breast
thermal images out of which 44 are normal and 28 are
abnormal. From the 75 breast thermal image 47 images are
used for testing and 25 are used for testing. Sample input
image is shown in the fig. 2.

![Image 2: Input Image From Visual Database](image)

B. Pre-Processing
The aim of pre-processing is an improvement of input
image data that suppress unwanted image data distortions or
enhance the some image features that are important for the
further processing. The Pre-processing that are considered
for this work includes Histogram equalisation, Contrast
stretching, Contrast Limited Adaptive histogram
Equalisation (CLAHE), Gamma Correction and
Decorrelation stretch.

Histogram Equalization is applied in order to improve
contrast of image without affecting the information
contained in it. Histogram equalization of the input image is
obtained using the equation (1).

\[
h(I) = \text{round} \left( \frac{cdf(i) - cdf_{\min}}{(mxl) - 1} \cdot X(L - 1) \right)
\]  

where \(i\) is the pixel intensity; \(cdf(i)\) is the cumulative
distribution function of \(i\); \(cdf_{\min}\) is the minimum non-zero
value of the \(cdf\); \(mxl\) represents the image size and \(L\) is the
number of grey levels used; \(h(I)\) is the equalised values of \(i\).

CLAHE reduces noise by partially reducing the local HE
and prevents the amplification of noise by clipping off those
pixels which are above the specified contrast limit and
distributed uniformly to other bins before applying histogram
equalisation. In Contrast stretching the intensity values of the
image are stretched to the entire range of values through
linear normalisation. Before the stretching, the image is
normalised using the upper and lower intensity limits to
obtain the new image with the intensity values within the
specified limits [5].

The input image is normalised prior to stretching using the
equation (2).

\[
I_N = \left( I - \text{Min} \right) \frac{\text{newMax} - \text{newMin}}{\text{Max} - \text{Min}} + \text{newMin}
\]  

The next pre-processing method used to enhance the input
image is the Gamma

![Image 1: Block diagram to identify the normal and
abnormal breast thermal images.](image)
correction that compensates the nonlinear luminance effect by mapping the luminance levels. The gamma value less than 1 produces good contrast that increases the accuracy of the segmentation. The decorrelation stretch removes the inter channel correlation in the input pixels to enhance the color differences in an input image.

The method maps the original colour values to the new set of colour values in wider range by transforming the intensity of each pixel into covariance matrix and then stretched to equalise the band variance. After equalizing the band variance it is again transformed to the original color bands.

The input thermal images are pre-processed using the above discussed method and the best suited method for the thermal breast images are identified using the performance metrics that includes the Absolute Mean Brightness Error (AMBE), Mean square Error (MSE), Peak Signal noise Ratio (PSNR), Enhancement measure Error(EME).

C. Image Segmentation

The important criteria are to distinguish the normality and abnormality in the breast tissues. Usually high intensity objects are more likely to be ill defined masses. The segmentation methods considered in this work are K means Clustering and Otsu thresholding. In K-means clustering the input preprocessed image are segmented into k clusters as given in the below algorithm

1. Number of k cluster is defined to label new region
2. Each data point is assigned to a nearest cluster by computing the sum of the squared distance between the data points and the cluster
3. Perform iteration until the data point doesn’t change the cluster

The algorithm iterates between steps one and two until a stopping criteria is met. The second segmentation method considered in this work is Otsu thresholding.

Otsu’s algorithm is a popular global thresholding technique that searches for a threshold that minimizes the intra-class variances of the segmented. If ‘I’ is an RGB image, a Karhunen-Loeve transform is performed on the Red, Green, Blue channels of the input image and the energy is compute. To the component with highest energy these segmentation is carried out.

The Otsu’s thresholding is global thresholding algorithm that involves the following steps.
1. Histogram of the input image is computed to obtain the probabilities of each intensity level.
2. Initial class probability (ωi) and class mean (μi) are set.
3. The maximum probability is set as threshold.
4. ωi and μi are updated and the difference between the class variance
5. The maximum difference value corresponds to set the desired threshold.

D. Feature Extraction

The Gray-Level Co Occurrence Matrix (GLCM), statistical features and structural features are extracted from the segmented region. The statistical features include Mean, Variance, Skewness, Standard deviation, Kurtosis. GLCM considers the spatial relationship of pixels in the gray level spatial dependence matrix. The GLCM features includes Contrast, Entropy Uniformity (also called Energy), and Homogeneity. The Shape based features such as area, perimeter, centroid, eccentricity, roundness are extracted from the segmented thermal images.

E. Feature Selection and classification

The significant extracted feature is selected using the student T Test method. The selected features are then given to the classifier and the performance of the classifier is analysed to identify the strong classifier to detect the normal and abnormal breast thermal images. The classifiers that are considered in this work includes Support Vector machine (SVM), KNN and Artificial Neural Network.

III. RESULTS

A. Image enhancement and segmentation

The following Fig.3 shows the sample and output images of various image preprocessing methods.

![Sample image](a) After Histogram Equalization c) Contrast Limited adaptive histogram equalization d) After contrast stretch (e) after decorrelation stretch (f) Gamma corrected result (for gamma value= 3) (g) Gamma corrected value (for gamma value=0.5)

Figure 3:

Performance metrics of the preprocessing methods are computed to identify the best suited method. Low value of MSE, High value of PSNR, High value of AMBE, High value of EME indicates best enhancement method. From Table I, it can be inferred that CLAHE and Contrast stretching are the best preprocessing methods.

The Fig. 4 shows the sample input image and segmented result using adaptive k-means clustering method and segmented result of Otsu thresholding method.
Detection of Breast Cancer from Thermography Images

Table I Image Enhancement Evaluation

| Method                | Controlled | Diseased |
|-----------------------|------------|----------|
| Histogram equalization |            |          |
| MSE                   | 1.96E+04   | 1.88E+04 |
| PSNR                  | 5.21E+00   | 5.37731  |
| EME                   | 1.17E+04   | 2.17E+04 |
| AMBE                  | 4.72E+03   | 112.164  |
| Contrast stretching   |            |          |
| MSE                   | 2.71E+03   | 6.68E+03 |
| PSNR                  | 1.39E+01   | 13.51076 |
| EME                   | 1.37E+01   | 3.113655 |
| AMBE                  | 8.70E+01   | 94.78103 |
| Decorrelation stretch |            |          |
| MSE                   | 6.28E+02   | 761.1885 |
| PSNR                  | 2.02E+01   | 19.45901 |
| EME                   | 7.60E+06   | 19.45901 |
| AMBE                  | 1.22E+02   | 122.0886 |
| CLAHE                 |            |          |
| MSE                   | 1.14E+03   | 3.05E+03 |
| PSNR                  | 1.76E+01   | 16.8627  |
| EME                   | 5.22E+00   | 3.677845 |
| AMBE                  | 1.42E+01   | 10.92693 |
| Gamma correction (γ=3) |            |          |
| MSE                   | 9.70E+01   | 95.43315 |
| PSNR                  | 9.70E+01   | 95.43315 |
| EME                   | 9.70E+01   | 95.43315 |
| AMBE                  | 9.56E+01   | 95.00866 |

Figure 4: (a) Sample image (b) Segmented result of Adaptive k-means clustering method (c) Segmented result for OTSU thresholding method

B. Feature evaluation

Before using these extracted features from segmented image as inputs for the classifiers in the classification step, they were analyzed using statistical student’s t test of the mean of control and diseased group samples to find significant features. As per this method, the features that have P value less than 0.05 were selected for the classification process. Analysis of Statistical features, Shape based features and GLCM features for controlled and diseased have been performed using student’s t test and P values are listed in the Table II, III and IV respectively.

Table II Statistical Feature Analysis For Controlled And Diseased

| Features     | Adaptive K means Clustering | Otsu’s thresholding |
|--------------|----------------------------|---------------------|
| Mean         | Control | Disease | P    | Control | Disease | P    |
| Variance     | 0.13    | 0.43    | 0.00 | 0.36    | 0.47    | 0.05 |
| Standard deviation | 0.11 | 0.08    | 0.02 | 0.16    | 0.18    | 0.19 |
| Skewness     | 1.04    | 0.57    | 0.25 | 0.01    | 1.49    | 0.05 |
| Kurtosis     | 0.15    | 0.63    | 0.05 | 24.69   | 0.51    | 0.37 |

Table III Shape Based Feature Analysis For Controlled And Diseased

| Features     | Adaptive K means Clustering | Otsu’s thresholding |
|--------------|----------------------------|---------------------|
| Mean         | Control | Disease | P    | Control | Disease | P    |
| Area         | 0.17    | 15.33   | 0.88 | 0.01    | 2372.11 | 0.05 |
| Perimeter    | 13.19   | 17.88   | 0.35 | 485.91  | 280.01  | 0.24 |
| Equivalent diameter | 3.42 | 5.41 | 0.04 | 0.69 | 0.30 | 0.01 |
| Eccentricity | 0.42    | 0.71    | 0.19 | 0.73    | 0.54    | 0.38 |
Table IV GLCM Feature Analysis For Controlled And Diseased

| Features                      | Adaptive K means Clustering | Otsu’s thresholding |
|-------------------------------|-----------------------------|---------------------|
|                              | Mean of controlled          | Mean of controlled  |
| Autocorrelation              | 0.23                        | 0.29                |
| Contrast                     | 2.35                        | 3.12                |
| Correlation                  | 0.88                        | 0.03                |
| Cluster Prominence           | 0.00                        | 0.90                |
| Cluster shade                | 0.01                        | 0.78                |
| Dissimilarity                | 0.41                        | 0.00                |
| Energy                       | 0.47                        | 0.48                |
| Entropy                      | 0.82                        | 0.90                |
| Homogeneity                  | 0.94                        | 0.94                |
| Maximum Probability          | 0.63                        | 0.56                |
| Sum of Square                | 0.24                        | 0.30                |
| Sum Average                  | 7.97                        | 8.12                |
| Sum Variance                 | 0.91                        | 0.87                |
| Sum Entropy                  | 0.78                        | 0.86                |
| Difference Variance          | 2.35                        | 3.00                |
| Difference Entropy           | 0.20                        | 0.24                |
| Information measure of Correlation 1 | -0.64           | -0.69                |

Based on the feature reduction technique discussed above, the most significant features where selected and given as an input for different classifiers to get better performance.

Mean, standard deviation, Equivalent diameter, Contrast, Dissimilarity, Maximum Probability Sum Entropy and Information measure of Correlation 1 are the significant features used to train the ANN, SVM and KNN for adaptive K-mean clustering method of segmentation.

Autocorrelation, Contrast, Equivalent diameter, Correlation, Cluster Prominence, Energy, Maximum Probability, Difference Variance and difference Entropy Homogeneity are the significant features used to train the ANN, SVM and KNN for Otsu threshold segmentation.

C. Performance measures of Classification Methods

The performance of the predictions was evaluated in terms of specificity, sensitivity, Classification accuracy, TP, TN, FP, FN, FPR, FNR, precision and error rate.

In a performance of classification test is measured using sensitivity and specificity. Fraction of positives that are correctly identified as such will be indicated by Sensitivity. The proportion of negatives that are correctly identified as such will be indicated by specificity.

Total number of correct predictions is indicated by Accuracy. Accuracy is viewed as balanced measure whereas sensitivity considers only positive cases and specificity considers with only negative cases. These performance metrics are calculated for different classifiers to detect the control and diseased segmented images which is given in Table V.

Table V Performance Measures Of Different Classifiers

| Performance measure | Adaptive K means Clustering | Otsu’s thresholding |
|---------------------|-----------------------------|---------------------|
|                     | SVM | ANN | KNN | SVM | ANN | KNN |
| TP                  | 8   | 8   | 8   | 6   | 4   | 6   |
| FP                  | 2   | 2   | 2   | 4   | 6   | 4   |
| FN                  | 5   | 0   | 5   | 0   | 0   | 0   |
| TN                  | 10  | 15  | 10  | 15  | 15  | 15  |
| Accuracy (%)        | 72  | 92  | 72  | 84  | 76  | 84  |
| Sensitivity (%)     | 61.54 | 100 | 61.54 | 100 | 100 | 100 |
| Specificity (%)     | 83.33 | 88.24 | 83.33 | 78.9 | 71.43 | 78.9 |
| FPR                 | 0.167 | 0.118 | 0.167 | 0.211 | 0.286 | 0.211 |
| FNR                 | 0.385 | 0.385 | 0.385 | 0 | 0 | 0 |
| Precision           | 0.8  | 0.8  | 0.8  | 0.6  | 0.4  | 0.6  |
| Error Rate          | 0.28 | 0.08 | 0.28 | 0.16 | 0.24 | 0.16 |

IV. DISCUSSION

The proposed algorithm is concentrated on finding appropriate classifier for the classification of breast images into normal or abnormal. The algorithm was tested with the DMR-IR Database images. The proposed classifiers are trained with 47 (29 normal, 18 abnormal) breast thermal images and tested with 25 breast thermal images (15 normal, 10 abnormal). From the results, it is inferred that ANN classifiers yields better classification accuracy of 92% and minimum error rate (0.08) in K-means segmentation method when compared with SVM, KNN classifiers. It is also witnessed that the classification accuracy of SVM and KNN (72%) are same for both normal and abnormal cases. SVM and KNN classifiers yields better classification accuracy of 84% and minimum error rate (0.16) in Otsu thresholding method when compared with ANN classifier. It is also observed that the classification accuracy of SVM and KNN are same for both normal and abnormal cases. The classification accuracy obtained for ANN in Otsu thresholding is 76% with 0.24 error rate.

Hence segmented image using in K-means segmentation method, ANN classifier are identified as the best methods for breast cancer diagnosis in this approach. Efficient feature selection algorithm such as Kernel F-score feature section, Principle Component Analysis, sequential search has to be implemented for improving classification accuracy. Also number of images for training has to be increased for improving classification accuracy.
Detection of Breast Cancer from Thermography Images

V. CONCLUSION

In this work, we have studied different image enhancement techniques to find out the suitable enhancement method for breast thermal image. CLAHE and Contrast stretching are identified as best methods. The preprocessed images are segmented using the adaptive K means clustering and Otsu method. The statistical features, GLCM and shape based features are extracted from the segmented image. Significant feature are selected using t-test method and then the classifiers namely the SVM, KNN, ANN, are trained with 47(29 normal, 18 abnormal) breast thermal images and tested with 25 breast thermal images (15 normal, 10 abnormal). From the results, it is inferred that ANN classifiers yields better classification accuracy of 92% and minimum error rate (0.08) in K-means segmentation method when compared with SVM, KNN classifiers. Further the work can be improved by analyzing feature selection algorithm such as Kernel F-score feature section, Principle Component Analysis, sequential search for improving classification accuracy.

REFERENCES

1. U.R. Acharya, E.Y.K. Ng, J.H. Tan, S.V. Sree, “Thermography based breast cancer detection using texture features and support vector machine(2012),” Journal of medical systems, Vol.3, No.3, pp. 1503-1510.
2. AmirEhsan Lashkari, Fatemeh Pak, Mohammad Firouzmand, “ Breast thermal images classification using optimal feature selectors and classifiers(2016),” Iranian Research Organization for Science and Technology (IROST), Vol. 2016, Is.7, pp. 237–248
3. Aya Hossam, Hany M. Harb, and H. M. Abd El Kader, “Performance Analysis of Breast Cancer Imaging Techniques (2017),” International Journal of Computer Science and Information Technology, Vol.15, No.5, pp.48-56.
4. Chinki Chandhok, Soni Chathurvedi, A.A Khurshad, “An approach to image segmentation using K-means clustering algorithm(2012),” UACEE International Journal of Artificial Intelligence and Neural Networks, Vol.2, Issue 2, ISSN 2250 – 3749.
5. Dinsha D, Manikandaprabu N, “Breast tumor segmentation and classification using SVM and Bayesian from thermogram images(2014),” Unique J. Eng. Adv. Sci, Vol. 2, No.2, pp. 147–151.
6. Hetal J. Vala, Astha Baxi, “A Review on Otsu Image Segmentation Algorithm(2013)” International Journal of Advanced Research in Computer Engineering & Technology (IARCE), Vol.2, Issue.2, February 2013, ISSN: 2278 – 1323.
7. Kanchanamani M, Varalakshmi Perumal, “Performance evaluation and comparative analysis of various machine learning techniques for diagnosis of breast cancer(2016),” Biomedical Research, Vol.27, No.3, pp. 623-631, ISSN 0970-938X
8. P. Kapoor, S.V.A.V. Prasad, S. Patni, “Automatic analysis of breast thermograms for tumor detection based on biostatistical feature extraction and ANN (2012),” International Journal of Emerging trends in Engineering and Development, Vol.7, No.2, pp.245-255.
9. J. Koay, C. Herry, and M. Frize, “Analysis of breast thermography with An artificial neural network (2004),” In Engineering in Medicine and Biology Society, IEMBS04, 26th Annual International Conference of the IEEE, Vol.1, pp. 1159-1162, IEEE.
10. M. Milošević, D. Jankovic, and A. Peulic, “Thermography based breast Cancer detection using texture features and minimum variance quantization (2014),” EXCLI Journal, Vol.13, pp. 1204-1215.
11. Mona A. S. Ali, Gehad Ismail Sayed, Tarek Gaber, Aboul Ella Hassanien, Vaclav Snasel, Lincoln F. Silva “Detection of breast abnormalities of thermograms based on a new segmentation method(2015), Proceedings of the Federated Conference on Computer Science and Information Systems, Vol.5 pp. 255–261 ACSIS.
12. Neeraj Bhargava, Anchal Kumawat, Ritu Bhargava, “Threshold and binarization for document image analysis using Otsu’s algorithm(2014),” International Journal of Computer Trends and Technology (IJCTT), Vol.17.No. 5, pp. 272-275, Nov 2014.
13. N. Scales, C. Herry, M. Frize, “Automated Image Segmentation for Breast Analysis Using Infrared Images (2004),” Annual International Conference of the IEEE Engineering in Medical and Biology Society, Vol.3, pp.1737-1740 IEEE.
14. N. Selvarasu, Alamelu Nachaiyappan, N.M. Nandhitha, “Image Processing Techniques and Neural Networks for Automated Cancer Analysis from Breast Thermographs(2012),” Indian Journal of Computer Science and Engineering(IJCSE), ISSN : 0976-5166, Vol. 3.No. 1, Feb-Mar 2012.
15. Sourav Pramanik, Debotosh Bhattacharjee, Mita N asipuri, “ Wavelet Based Thermogram Analysis for Breast Cancer Detection(2015)” International Symposium on Advanced Computing and Communication (ISACC)Vol.15, Sep 2015.
16. H. Qi, and F. H. Jonathan, “Asymmetry analysis using automatic segmentation and classification for breast cancer detection in thermograms (2001),” 23rd Annual International IEEE Conference on Engineering in Medicine and Biology Society, Vol. 3, pp.2866-2869.

AUTHORS PROFILE

Suganthi K. is working as Associate professor in the Department of Biomedical Engineering, SSN college of Engineering. Her Research area includes medical signal and image processing, embedded system design, Biomedical instrumentation

Nirmala K. is working as Associate professor in the Department of Biomedical Engineering, SSN college of Engineering. He had completed her Ph.D in Anna University, Chennai. Her Research area includes medical image processing, cryprography,

Shaalu Sree B.N is Under Graduate scholar in the Department of Biomedical Engineering, SSN college of Engineering.

Pawithra S is Under Graduate scholar in the Department of Biomedical Engineering, SSN college of Engineering

Yashvantha N is UG scholar in the Department of Biomedical Engineering, SSN college of Engineering.