ISSN: 2321-7782 (Online)
Volume 3, Issue 6, June 2015
International Journal of Advance Research in Computer Science and Management Studies
Research Article / Survey Paper / Case Study
Available online at: www.ijarcsms.com

Classification of Mammogram Images using GLCM and Trace Transform Functionals

Kamalu B.Nair[1]
Electronics and Communication Engineering
Amal Jyothi College of Engineering
Kerala, India

Binoshi Samuvel[2]
Electronics and Communication Engineering
M.Tech, CUSAT
Kerala, India

Abstract: Mammography is one of the first diagnostic tests to prescreen breast cancer. Early detection of breast cancer has been known to improve recovery rates to a great extent. It is difficult for radiologists to identify the masses on a mammogram because they are surrounded by complicated tissues. Computer-aided detection (CADe) systems have been developed to aid radiologists in detecting mammographic lesions which may indicate the presence of breast cancer. Trace transform, which is a generalization of the radon transform and GLCM(grey-level co-occurrence matrix) together has been used to extract the features .GLCM compute four cooccurence matrices with one pixel distance in four directions: left diagonal, right diagonal, vertical and horizontal.Four statistics can be calculated by describing the image texture i.e energy , contrast , correlation , homogenity are four texture features. Gaussian classifier is used in the classification of mammogram images into normal, abnormal, benign and malignant classes.

Keywords: Cancer, mammogram, texture, trace transform.

I. INTRODUCTION

Breast cancer death rates are higher than that of any other cancers for women in the U.S. Approximately 39,520 women in the U.S died from breast cancer in 2011, although death rates have been decreasing. These decreases are likely the result of treatment advances, increased awareness, and early detection. Screening mammography is one of the most effective techniques for the early detection of breast cancer.

A radiologist typically examines a mammogram to check for signs of cancer. Computer-aided detection (CADe) system prompts the radiologist to re-examine the films. When using a CADe system with mammography, a radiologist still reads the mammogram, but a computer program also evaluates the mammogram and highlights suspicious regions for the radiologist to review. Finally, the radiologist identifies true areas of concern before making a final diagnosis. The CADe system in screening mammography serves as a second opinion that calls attention to abnormalities and avoiding unnecessary biopsies [1].

Masses may also have different shapes and margins, may differ in size, location, and orientation, and may have different backgrounds. Masses with irregular shapes are highly suspicious for breast cancer. Breast masses are more difficult to identify because of the abundant appearances and ambiguous margins compared to calcifications [3]. Thus, mass detection continues to challenge both radiologists and CADe systems. Researchers have developed many schemes for mammographic mass detection.

Classification systems [2] can help in minimizing possible errors and also can provide instant examination of medical data in a relatively short time and in a fairly detailed manner. CAD systems combine a variety of techniques from the fields of artificial intelligence and digital image processing along with the domain knowledge from medical experts to support cancer detection. CAD software help radiologists to better detect masses and microcalcifications in mammograms, and hence can improve the accuracy of mammography and also reduce the subjectivity associated with the manual interpretation process. In

CAD software, the mammograms are first enhanced using standard image enhancement methods mainly to sharpen the boundaries of the region of interest (ROI) and to increase the contrast between the ROI and the nearby normal tissue.

GLCM (grey-level co-occurrence matrix) are used for feature extraction. These methods are more accurate compared to trace functionals.They describe the textures by a matrix of pair-gray level appearing probabilities[6].They compute four cooccurence matrices with one pixel distance in four directions: left diagonal, right diagonal, vertical and horizontal. Four statistics can be calculated by describing the image texture i.e energy, contrast, correlation, homogenity are the four texture features.

II. Proposed System

Our main goal is to detect breast cancer and classify the images as normal and abnormal mamamogram images. The flowchart of the proposed system is given in Figure1. The mammogram images are selected from the dataset and feature extraction is done using Trace transform funtionals and GLCM (grey-level co-ocurrence matrix) as the feature extraction methods.Trace transform is a generalisation of radon transform. GLCM describe the textures by a matrix of pair-grey level appearing probabilities and computes four cooccurence matrices with one pixel distance in four directions: left diagonal, right diagonal, vertical and horizontal.Four statistics are calculated by describing the image texture i.e energy, contrast, correlation and homogenity are the four texture features.Gaussian classifiers is used for classification of mammogram images into normal, abnormal, benign and malignant classes.

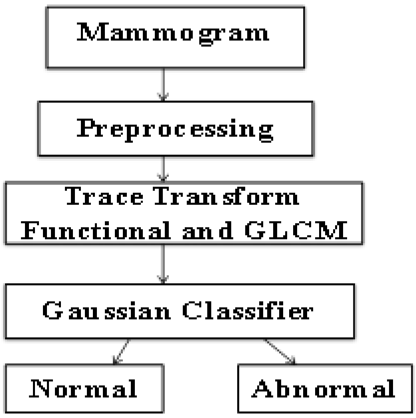

Figure1: Flowchart

III. Feature Extraction

Trace transform functionals were used for feature extraction. Trace transforms are a generalization of the Radon transform where the transform calculates functionals of the image along lines tracing through its pixels. The trace transform works by transforming the original image into a mapped image, which is a 2-D function, based on a set of parameters that characterize each line criss crossing its domain. An example of a set of trace transforms for normal, benign, and malignant images is shown in figure3, respectively. The transformed image is a 2-D function, which can be further reduced to a string of 1-D arrays by using another set of functionals called the diametric functionals .Finally, from these strings of values, a third functional called the circus function is calculated to yield a single number, which is referred to as the triple feature. The triple feature is unique for every image, and various combinations of functionals produce different triple features for the same image, allowing the user to generate several features for the same image. The triple feature is produced using functionals that are usually invariant to rotation, translation, and scaling.

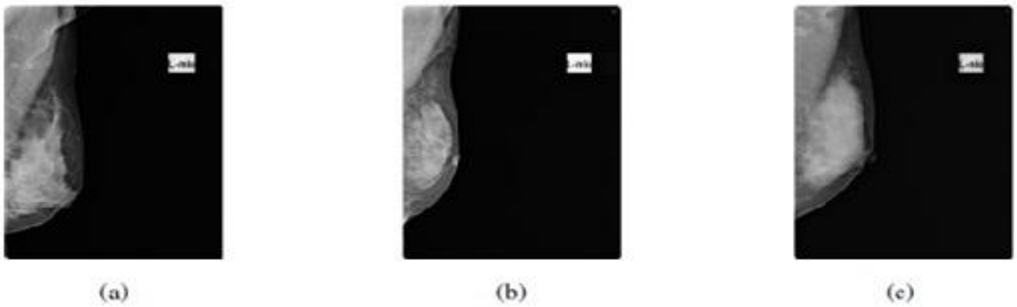

Figure 2: Sample mammogram images (a) normal,(b)benign and (c) malignant

Generalisation of Radon Transform

Similar to the trace transform using the functionals, the Radon transform uses a specific trace functional which is the integral of the function. This proves that Radon transform is just a generalization of the trace transform [3]. The trace transform has twice the computational complexity of the Radon transform. This is due to the fact that the integral is a functional that retains the same absolute value which ever way the tracing line is traversed and each line has to be considered only once. But the trace functional lines have to be traversed both forward and backward in order to account for the variability arising out of the variation in functional values in each direction. In order to exploit all the advantages of trace transform functionals, one must choose only those functionals that are invariant to rotation, translation, and scaling of the image [3].

Triple feature extraction

The basis of trace transform feature extraction is the assumption that objects are subjected to linear distortions in the form of rotation, scaling, and translation. In other words, the image remains the same but is viewed from a linearly distorted coordinate system. For instance, if the original coordinate system of the image is C_1 and the distorted coordinate system is C_2, an image F(x, y) seen from C_2 will be the same as being seen from C_1 as long as the transformation is linear because linear transformations preserve lines. Consider this image F(x, y) to be scanned with lines in all directions. The trace transform is a function g, where T is referred to as the trace functional. If L (C_1; φ, p, t) is a line in the coordinate system C_1, then:-

$$g(F;C_1;\Phi;p)=T(F(C_1;\Phi;p;t))$$

where F (C1; φ, p, t) means the values of the image function along the chosen line. This functional results in a 2-D function of the variables. A triple feature is defined with the help of two more functionals referred to as diametrical and circus functions. Diametric is a functional of the trace transform function when it is considered as a function of the length of the normal to the line. Circus is a functional operating on the orientation variable after the previous two operations have been performed. So, the triple feature Π is defined as :-

$$\Pi\ (F,C_1) = \Phi\ (p\ (T\ (F\ (C_1;\ \Phi;\ p;\ t))))$$

S.No	Trace Functional
1	$\sum_{i=1}^{N} x_i$
2	$\sum_{i=1}^{N} ix_i$
3	$\frac{1}{N}\sqrt{\sum_{i=1}^{N}(x_i-\hat{x})^2}$
4	$\sqrt{\sum_{i=1}^{N} x_i^2}$

Table I: Trace Functional T, xi is the grey value of the image at point i along the tracing line and N is the total number of points.

S.No	Diametric Functional
1	$Max_{i=1}^{N} x_i$
2	$Min_{i=1}^{N} x_i$
3	$\sqrt{\sum_{i=1}^{N} x_i^2}$
4	$\frac{\sum_{i=1}^{N} i x_i}{\sum_{i=1}^{N} x_i}$

Table II: Diametric Functional P used for analysis, xi is the value of the trace transform at row i along the column to which functional is applied and N is the total number of rows of the trace transform.

S.No	Circus Functional
1	$\sum_{i=1}^{N-1} \lvert x_{i+1} - x_i \rvert^2$
2	$\sum_{i=1}^{N-1} \lvert x_{i+1} - x_i \rvert$
3	$\sqrt{\sum_{i=1}^{N} x_i^2}$
4	$\sum_{i=1}^{N} x_i$

Table III: Circus FunctionalΦ used for analysis, xi refers to the value of the circus function at angle i and N is the total number of columns of the trace transform.

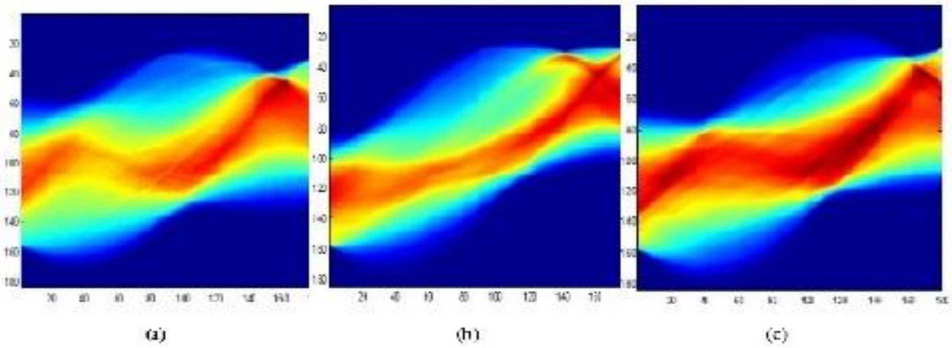

Figure 3: Sample Trace Transform images (a) normal, (b) abnormal and (c) malignant

GLCM (Grey- level Co-occurrence Matrix)

The second step is to extract the statistical features such as contrast, correlation, energy and homogeneity is to compute the GLCM matrix of the input texture image [6].The GLCM is normalized so that the sum of its elements is equal to 1. Each element (i,j) in the normalized GLCM is the joint probability occurrence of pixel pairs with a defined spatial relationship having gray level values i and j in the image. Let us consider p is the normalized GLCM of the input texture image. Contrast is a measure of the intensity contrast between a pixel and its neighbor over the whole image.

$$\text{Contrast}=\sum |i-j|\, p(i,j)^2$$

$$\text{Correlation}=\sum \frac{(i-\mu i)(j-\mu j)p(i,j)}{\sigma i \sigma j}$$

The energy is the sum squared element in the normalized GLCM and the homogenity is a value that measures the closeness of the distribution of elements in the GLCM to the GLCM diagonal.

$$\text{Energy}=\sum p(i,j)^2$$

$$\text{Homogenity}=\sum \frac{p(i,j)}{1+|i-j|}$$

IV. Feature Selection, Ranking And Classification

The problem of over training a classifier also becomes pronounced where there are a large number of features. An example for this is the present case, where a feature space with more than 1000 dimensions owing to the several possible permutations that can be built from the three different functionals. To avoid the problem of over training, the number of degrees of freedom in the function has to be decreased, and for this, the number of features has to be decreased. This is done by feature ranking and selection.

The feature selection procedure is carried out by selecting a class separability criterion C and computing its value for all available featuresx_k, k= 1, 2, . . . , m. The C value is then ranked in descending order. The one with the best C value is named asx_{i1}. The cross-correlation coefficient between x_{i1} and the remaining features are calculated to select the second best feature. If α_1 and $\alpha_2$are the two features in consideration with C being the class separability criterion and ρ representing the position of the features in the original set, the feature x_{i2} for which i_2= argmax j $\{\alpha_1 C\,(j) - \alpha_2|\rho_{i_1 j}\,|\}$, for all j≠$i_1$ is calculated to find the next best feature the procedure is carried on further for all values of k and the final GLCM texture features i.e energy, contrast, correlation, homogenity are used to rank the best features.

Gaussian classifier

It is a probabilistic model that is usually used in the case of data that has a sub population within a universal population. The advantages of using a GMM is that it is computationally tractable and often provides good results in classification.

V. Experimental Results

The performance of the proposed system is evaluated using 109 normal images, 113 abnormal images, 62 benign and 51 malignant images. Gaussian classifiers classify the images into normal, abnormal, benign and malignant classes.Initially the testing image is selected and is used for feature extraction using trace transform functional and GLCM. Then the classification of the mammogram is done using gaussian classifier by computing the mean, variance and the covariance. Simulation is done using Matlab and the GUI (graphical user interface) is the enviornment for selecting the images and the different stages.

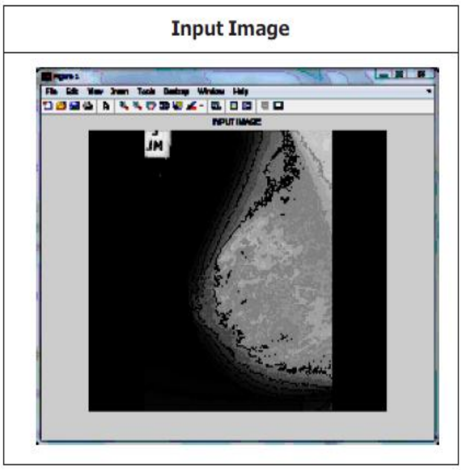

Figure 4: Input mammogram image

Initially the input image is selected from the set of testing images.Then feature extraction is done using trace transform functional and GLCM.Trace transform functional is a generalisation of radon transform. GLCM computes four cooccurence matrices with one pixel distance in four directions: left diagonal, right diagonal, vertical and horizontal.Four statistics can be calculated by describing the image texture i.e energy, contrast, correlation, homogenity are the four texture features.

Gaussian classifiers classify the images into normal, abnormal, benign and malignant classes by computing the mean, variance and covariance. The calculations are done separately. The combination of all these parameters and the dataset is used in the generation of set of values for normal and abnormal classification of mammograms. The values generated are saved and stored in seperate workspaces. Hence the process of classification is complete and the different stages of the mammogram images are obtained.

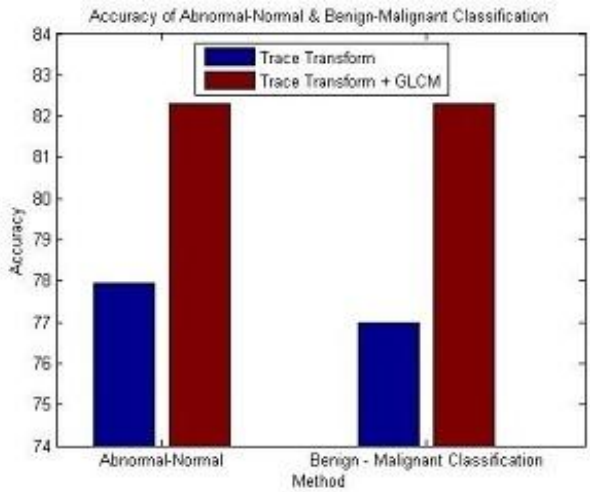

Figure 5: Classification Accuracy

VI. CONCLUSION

A novel framework used in mammographic image analysis has been proposed. Trace transform along with GLCM provides 85% classification accuracy and provides with better performance compared to trace transform functional alone .Results show that the proposed system achieves satisfactory detection performance.GLCM is used in non-seismic purposes such as remote sensing.GLCM techniques are mainly used in pattern recognition by using various statistical techniques.The results show these feature extraction methods have high discrimination accuracy, requires less computation time and hence efficiently used for real time pattern recognition applications.

ACKNOWLEDGMENT

The authors would like to thanks to Matha hospital, for providing the dataset and like to thank our Institution for their constant support and encouragement. Also I would like to thank Dr.Jaipal Matha hospital, Kottayam for their valuable information.

References

1. Karthikeyan Ganeshan, K.Thomas Abraham, "One class classification of mammograms using trace transform functionals", IEEE transactions on instrumentation and measurement vol.63.no:2, February 2014.
2. Zhang and Noriyasu Homma, "A Hybrid Image Filtering Method for Computer-Aided Detection of Microcalcification Clusters in Mammograms", Hindawi Publishing Corporation Journal of Medical Engineering Volume 2013.
3. S Mohan Kumar and G. Balakrishnan, "Mass Classification Analysis in Digital Mammogram using DWT and SNE", International conference on Communication and Signal Processing, April 3-5, 2013, India.
4. Ivan and Blagojce, "Comparison of Feature Extraction Algorithms for Mammography Images", 4th International Conference on Image and Signal Processing 2011.
5. T.A. Sangeetha and A. Saradha, "An Efficient Way to Enhance Mammogram Image in Transformation Domain", International Journal of Computer Applications Volume 60– No.2, Dec 2012.
6. S Murali and P Satheesh Kumar, "Multiclass Classification of Mammogram Images with GLCM features", International Journal on Engineering Research and Science and Technology, Vol. 4, No. 1, February 2015.
7. A. Suresh and K.L. Shunmuganathan , "Feature Fusion Technique for Colour Texture Classification System based on Grey level Co-occurrence Matrix", International Journal of Computer Science Volume 8,December 2012.

AUTHOR(S) PROFILE

Ms. Kamalu B.Nair, pursuing M.Tech Degree in Communication Engineering Degree in at Amal Jyothi College of Engineering, Kottayam, Kerala, India. She has obtained her Bachelor Degree from Mahatma GandhiUniversity, Kerala, India.

Mr. Binoshi Samuvel, M. Tech works as Assistant Professor at Amal Jyothi College of Engineering, Kottayam, Kerala, India. His areas of interest include Digital Signal Processing, Wavelets, Multirate Signal Processing and Adaptive Signal Processing

ISSN: 2321-7782 (Online)