

A New Approach to Determine the Classification of Mammographic Image Using K-Means Clustering Algorithm

S.Julian Savari Antony¹, Dr.S.Ravi²

¹Research Scholar, Department of Electronics and Communication Engineering, Shri J.J.T. University, Jhunjhunu, Rajasthan, India, ²Faculty of Electrical Engineering, Selvam College of Technology, Namakkal, Tamil Nadu, India
Email:savari.sm@gmail.com

ABSTRACT

Breast cancer is one of the most emergent disease in women. Image classification is a supporting for medical system as well as the difficult task also. Mammographic images contain various features like space, distance, circumscribed masses and microcalcification. This paper addresses on classifying mammographic images created on the Region of Interest (ROI). The images obtained may be blurred due to the variation in illumination, intensity and contrast which cannot be directly processed. It will be discussed two stages for classify the image. At first stage, the quality of image is increased using histogram equalization method which normalizes the image. At second stage, the intensity features are computed from the image of shape features, region features such as 'Solidity', 'Eccentricity', 'Convex Area', 'Orientation', 'Perimeter', 'Major Axis Length', 'Minor Axis Length' which are extracted to compute volumetric values. The proposed classification method is reduced by the false positive result arise in mammogram classification and overcome missing features while using single feature mammogram classification. Therefore, the quality of the image classification is a final aim of this paper that is improved. The result shows that the proposed method has increased the classification accuracy.

Keywords : Mammogram image; Region of interest; Feature extraction; K means clusterin; Time complexity

1 INTRODUCTION

Breast cancer is one of the major causes of death among women. Small clusters of microcalcifications looking as collection of white spots on mammograms show an early warning of breast cancer. Primary deterrence seems impossible since the causes of this disease is still remain unknown. An improvement of early diagnostic techniques is critical for women's quality of life. Mammography is the main test used for screening and early diagnosis. Early detection performed on X-ray mammography is the key to improve breast cancer prognosis. In order to increase radiologist's diagnostic performance, several computer-aided diagnosis (CAD) schemes have been developed to improve the detection of primary signatures of this disease: masses and micro calcifications. Masses are space-occupying lesions, described by their shapes, margins, and denseness properties. A benign neoplasm is smoothly marginated, whereas a malignancy is characterized by an indistinct border that becomes more speculated with time. Because of the slight differences in X-ray attenuation between masses and benign glandular tissue, they appear with low contrast and often very blurred. Micro calcifications are tiny deposits of calcium that appear as small bright spots in the mammogram. Clustering is a division of data into groups of related objects. Every group, called a cluster, consists of objects that are similar between themselves and different compared to objects of other groups. Cluster analysis is a main task of Data Mining. It divides the datasets into several meaningful clusters to reflect the data sets' natural structure. Cluster

Copyright © 2015 SciResPub.

is aggregation of data objects with common characteristics based on the measurement of some kind of data. There are several commonly used k-means algorithm, hierarchical clustering algorithm, self-organizing maps algorithm, expectation maximization clustering algorithm and Fuzzy C means. All these algorithms are compared according to the following factors: size of dataset, number of clusters, type of dataset, intensity - based segmentation and type of software used. Some conclusions that are extracted belong to the performance, quality, and accuracy of the clustering algorithms [1],[2]. In [3] aims to develop a software that can define the stage of breast cancer based on the size of the cancerous tissue. Steps of the research consist of mammogram image acquisition, determining the Region of Interest, using Region growing segmentation method [12], measuring the area of suspected cancer, and determine the stage classification of the area on the mammogram image by using Sample K-Means Clustering method.

In this paper two types of classifications support vector machine and linear discriminant analysis are used to analyse the mammographic images. The two classification methods are using the image pre-processing in wavelet decomposition and image enhancement [4]. There are several stages to breast image processing. The first stage, breast image acquisition through mammography. The next stages are pre-processing image, segmentation, feature extraction, feature selection and classification [5]. With technique digital mammography, char-

acteristics calcification, circumscribed, speculated and other ill-defined masses can be diagnosed [6].The paper [7] has following work plans: at first the quality of image is increased using histogram equalization method. At second, the intensity features are computed from the image. The classification of mammograms with benign, malignant and normal tissues using independent component has been studied in [8].Optimal set of features selected by Genetic algorithm are fed as input to Adaptive Neuro fuzzy inference system for classification of images into normal, suspect and abnormal categories [9].

Detection of edges in an image is a very important step towards understanding image features. Since edges often occur at image locations representing object boundaries, edge detection is extensively used in image segmentation when images are divided into areas corresponding to different objects. This can be used specifically for enhancing the tumor or cancer area in mammographic images. Different methods are available for edge detection like Roberts, Sobel, Prewitt, Kirsch and Laplacian of Gaussian edge operators [10], [13]. Region Growing is a procedure that classifies the pixels or sub-regions into larger regions based on predefined criteria. The approach basically starts from the beginning of the set of points, then the area is enlarged by adding each neighboring pixel point that has properties similar to those points (for example the range of intensity or color specification). The selection of similar criteria, in addition to depending on the problem at hand, also depends on the type of image data available, for examples of descriptors include moment and texture [11].

The aim of this paper is to propose a new approach to determine the classification of Mammographic Image Using K-Means Clustering Algorithm. This work is organized as follows. In Section 2, it present the proposed method includes the process of classification. Next, in Section 3, the results are shown. Finally, Section 4 presents some concluding observations.

2 PROPOSED METHOD

The proposed method has following work plans in figure. 1 system Architecture. The quality of image is increased using histogram equalization method which normalizes the image. Then the intensity features are computed from the image and then shape features, region features like 'Solidity', 'Eccentricity', 'Convex Area', 'Orientation', 'Perimeter', 'Major Axis Length', 'Minor Axis Length' are extracted to compute volumetric values. Computed feature set is used to classify the image using k-means clustering.

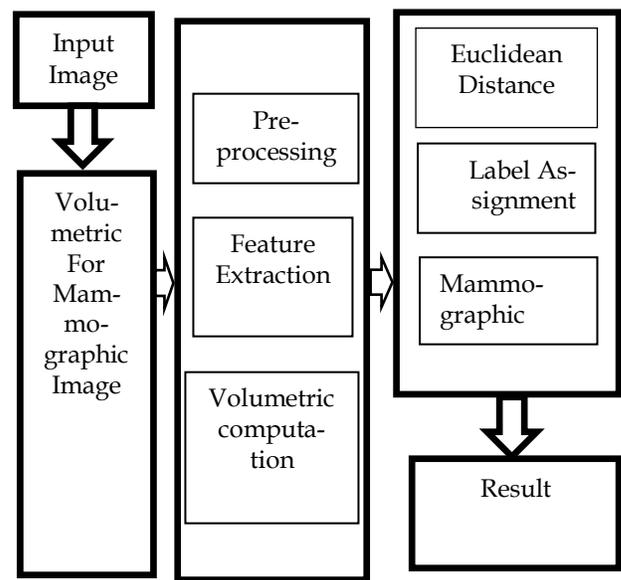


Figure.1. System Architecture

2.1 A. Pre-Processing Final Stage

The process of removing noise present in the image will be removed in this stage. It shows in figure.2 input image which is applied with the Gabor filter to remove the noisy signals present in the image. The noise removed image is applied with the histogram equalization techniques in figure.3 which normalize the intensity values of the image. The pre-processed image has uniform features with higher intensity value persists. A 2-D Gabor function is a Gaussian modulated by a sinusoid. It is a non-orthogonal wavelet. Gabor filters exhibits the properties as the elementary functions are suitable for modelling simple cells in visual cortex and gives optimal joint resolution in both space and frequency, suggesting simultaneously analysis in both domains. The definition of complex Gabor filter is defined as the product of a Gaussian kernel with a complex sinusoid.

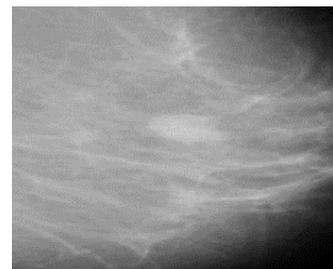


Figure.2. Original image

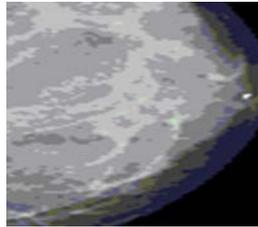


Figure.3.Histogram equalization image

2.3 Feature Extraction

The feature extraction has provide to accurate recognition and feature patterns. In particular, the significant features must be encoded in this experimentation and used to extract the intensity features which is the Gabor filter with low, high frequencies and four different orientations. Three low frequencies and three high frequencies with four orientations give 12 combinations of Gabor filter. In order to extract the region properties have implemented by Gaussian filter. Then the filtered image is converted into binary image that is identified by the regions and the features named solidity, eccentricity, convex area, orientation, perimeter, major axis length, and minor axis length are extracted. The extracted features are stored in a feature vector which will be used in further analysis.

2.4 Volumetric Estimation

From computed region and shape features the volume of the shape is calculated using multiple points of the region. Using the distance metrics calculate the overall volume of the component which is calculated the different co-ordinate points. The boundary points are identified and split the points into simple shapes less than polygon. Based on the separated multi-dimensional space, for each shape volume will be compute.

2.5 Mammographic Classifier

It classifies the mammographic image and generated feature vectors which are evaluated by each feature such as intensity value, density measure, solidity, eccentricity, convex area, orientation, perimeter, major axis length, and minor axis length. The feature vector has estimated to number of intensity values, density measure and according to the feature vector is the size of the pixels covered under the shape or region. Its mean value of intensity and density is computed based on the pixel values present in the region. Rests of the features are left as it is to calculate the Euclidean distance by the k-means clustering.

The mammographic classifier computes the Euclidean distance iteratively to evaluate the relative measure between the feature vectors present in the training and testing set. Based on the mean values computed with different class be-

tween feature vectors the classifier assigns a label for the submitted test image.

3 RESULTS AND DISCUSSION

The proposed methodology has been evaluated with various data sets of mammographic images. The proposed method uses different data set and various numbers of classes namely classification, circumscribed masses, speculated masses, ill-defined masses, architectural distortion and asymmetry. Number of samples has been used for each class. The following Table I shows the data set used for the evaluation of the proposed method. The Department of Defense (DoD) data base for breast cancer research program of Stanford School of medicine, California and Mammographic Image Analysis Society (MIAS) data set and Digital Data set for Screening Mammography (DDSM) for the evaluation of the proposed method. The proposed method uses less number of samples, i.e. less than 10% of training samples. Also it takes less time for training and testing, because unlike other methods, only few stages are used to identify the mammogram and image classification

TABLE 1
 SHOWS THE DETAILS OF DATASET USED

Database	Number of samples	Number of test- ing images
DoD BCRP	750	180
MIAS	322	115
DDSM	2640	43

It is shown in Figure.4. Process Flow of K-Means Clustering Algorithm which is focused into the feature vectors. The database is applied to the K-means clustering which is produced in Figure.5 shows the identified interest points and it is clear that the region marked with red colour represents the presence of cancer cells.

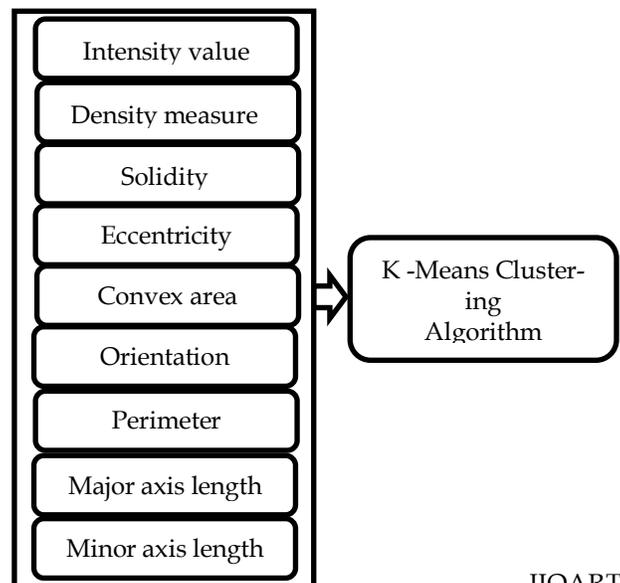


Figure.4. Process Flow of K-Means Clustering Algorithm

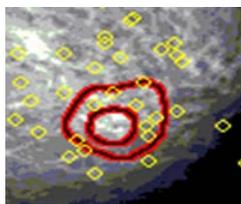


Figure.5. Identified interest points

The Table 2. shows the percentage of classification accuracy against five different algorithms. It clearly depicts that the higher classification accuracy of 99% is attained through the proposed K means clustering has produced higher classification accuracy than the existing methods taken for comparison. It is clear that the proposed method produces higher accuracy at all type of class combinations.

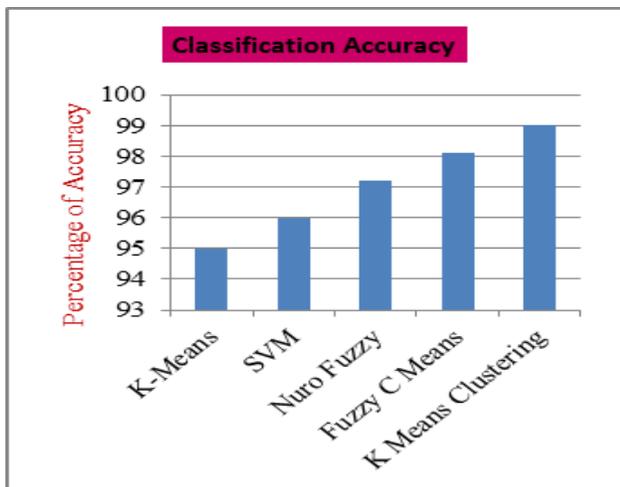
TABLE 2
 EVALUATION OF CLASSIFICATION ACCURACY

Algorithm	Classification Accuracy
K-Means	95
SVM	96
Neuro Fuzzy	97.2
Fuzzy C Means	98.1
K Means Clustering	99



Graph2: shows the time complexity of different methods

The graph2 shows the time complexity produced by different algorithms. It is clear that the proposed multi-variant approach has produced less time complexity which shows the more classification accuracy.



Graph1: shows the classification accuracy.

The graph1 shows the classification accuracy of different algorithms and it is clear that the proposed multi-variant approach has produced higher classification accuracy compare to other methods.

4 CONCLUSION

This proposed a new approach to determine the classification of mammographic image using k-means clustering algorithm which uses different features of the image like intensity values, shape and region features and density features to compute the feature vector. They have computed mean values of the intensity values of the pixels in the region extracted to compute the intensity mean value and the density measure is also computed in the similar fashion. The region metric is computed using the extracted region values and it has seven different features hidden. Based on the computed feature vectors the k-means clustering is used to identify the class of the input image. The proposed system produces good results and reduces the space and time complexity. It has produced classification accuracy up to 99 % which is more than other methodologies in this era. In future, we will consider other clustering models for feature extraction in order to improve the classification rate.

REFERENCES

[1] Osama Abu Abbas, Jordan, "Comparisons Between Data Clustering Algorithms," The International Arab Journal of Information Technology, vol. 5, no. 3, pp.320-326, Jul. 2008.

- [2] S.Saheb Basha, Dr.K.Satya Prasad," Automatic Detection Of Breast Cancer Mass in Mammograms using Morphological Operators and fuzzy c -means clustering", Journal of Theoretical and Applied Information Technology, Vol.7. No.6,2005.
- [3] Karmilasari, Suryarini Widodo, Matrisya Hermita, Lussiana ETP," Sample K-Means Clustering Method for Determining the Stage of Breast Cancer Malignancy Based on Cancer Size on Mammogram Image Basis", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 5, No. 3, 2014
- [4] S.Julian Savari Antony and S.Ravi, "Development of Efficient Image Quarrying Technique for Mammographic Image Classification to Detect Breast Cancer with Supervised Learning Algorithm", Proceedings of the IEEE Xplore on ICACCS, pp.1-7, Dec. 2013.
- [5] Bozek, J., Mustra, M., Delac, K., and Grgic, M., "A Survey of Image Processing Algorithms in Digital Mammography", Multimedia Signal Processing and Communications Studies in Computational Intelligence Volume 231, 2009, pp 631-657.
- [6] Yasmin, M., Sharif, M., and Mohsin, S., "Survey Paper on Diagnosis of Breast Cancer Using Image Processing Technique", Research Journal of Recent Sciences, Vol. 2(10), 88-98, October 2013.
- [7] S.Julian Savari Antony and S.Ravi, "Development of Efficient Image Quarrying Technique for Mammographic Image Classification to Detect Breast Cancer with Supervised Learning Algorithm", Proceedings of the IEEE Xplore on ICACCS, pp.1-7, Dec. 2013.
- [8] R.M.Haarlick, "Statistical and structural approaches to texture", Proceedings of the IEEE, vol.67, no.5, pp.786-804, 1979.
- [9] S.Julian Savari Antony,"Detected Breast Cancer on Mammographic Image Classification Using Fuzzy C-Means Algorithm" International Journal of Innovations in Engineering and Technology (IJJET), Volume 4 Issue 4 December 2014.
- [10] Maitra, I.K., Nag S., Bandyopadhyay S.K., "A Novel Edge Detection Algorithm for Digital Mammogram", International Journal of Information and Communication Technology Research, Vol 2 No.2, February 2012.
- [11] Kamdi, S," Image Segmentation and Region Growing Algorithm", International Journal of Computer Technology and Electronics Engineering (IJCTEE), Vol 2 Issue no.1, October 2011.
- [12] Priya, D.S., and Sarojini, B., "Breast Cancer Detection in Mammogram Images Using Region- Growing and Contour Based Segmentation Techniques", International Journal of Computer & Organization Trends, Vol3 Issue 8, September 2013.
- [13] S. Julian Savari Antony, "Improvement of Efficient Volume Estimation Mammographic Image Classification Based on wavelet Analysis", Vol. - III Issue - IV ISSN 2277 - 5730.