Detected Breast Cancer on Mammographic Image Classification Using Fuzzy C-Means Algorithm

S.Julian Savari Antony
Research Scholar, Department of Electronics and Communication Engineering
Shri J.J.T. University, Jhunjhunu, Rajasthan, India

Abstract- Breast cancer is one of the most common tumour in women. It is a foremost cause of death in the world. A proper screening procedure can help an early diagnosis of the tumor so that reducing the death risk. A suitable computer aided detection system can help the radiologist to sense many subtle signs, normally missed during the showing phase, submitting to the radiologist's attention those areas that could comprehend an irregularity. The proposed method has following work plans: at first 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 and then shape features, region features which are extracted to compute volumetric values. Computed feature set is used to classify the image using fuzzy c means clustering which reduces false positive result arise in mammogram classification and overcome missing features while using single feature mammogram classification. Experimental results show that, when compared to several other methods fuzzy c means shows 98.1% microcalcification detection in mammograms

Keywords – Mammography, Breast cancer, Euclidean Distance, Fuzzy C means,

I. INTRODUCTION

Mammography Cancer causes 1 in 8 deaths worldwide and is rapidly becoming a global pandemic. According to the International Agency for Research on Cancer, there were 12.7 million new cancer cases up to 2008. If the rates don't change, the global cancer burden is expected to be nearly doubled (i.e.,) 21.4 million cases and 13.5 million deaths by 2030. According to the World Health Organization (WHO), the toll of cancer and other chronic diseases is greater in low and middle income countries where publics develop chronic diseases at younger ages who suffer longer – often with preventable complications – and die sooner than those in high-income countries [1].

Mammographic image classification using Histogram Intersection [2], proposed using histogram intersection for mammographic image classification. First, we use the bag-of-words model for image representation, which captures the texture information by collecting local patch statistics. Then the normalized histogram intersection (HI) is proposed as a similarity measure with the K-nearest neighbor (KNN) classifier. Furthermore, by taking advantage of the fact that HI forms a Mercer kernel, we combine HI with support vector machines (SVM), which further improves the classification performance. In Mammogram Image Classification Using Fractal Properties [5], given a mammogram image, first all the artifacts are removed and the salient features are extracted such as Fractal Dimension (FD) and Fractal Signature (FS). These features provide good descriptive values of the region. Second, a trainable multilayer feed forward neural network has been designed for the classification purposes and we compared the classification test results with K-Means.

In [3], Gacsadi proposes a directional features based automatic tumor classification of mammogram images. In this patches around tumors are manually extracted to segment the abnormal areas from the remaining of the image, considered as background. The mammogram images are filtered using Gabor wavelets and directional features are extracted at different orientation and frequencies. Principal Component Analysis is employed to reduce the dimension of filtered and unfiltered high-dimensional data. Proximal Support Vector Machines are used to final classify the data. Superior mammogram image classification performance is attained when Gabor features are extracted instead of using original mammogram images. The robustness of Gabor features for digital mammogram images distorted by Poisson noise with different intensity levels is also addressed.

Associative Classification of Mammograms Using Weighted Rules [4], presents a novel method for the classification of mammograms using a unique weighted association rule based classifier. Images are pre-processed to reveal regions of interest. Texture components are extracted from segmented parts of the image and discretized for rule discovery. Association rules are derived between various texture components extracted from segments of images, and employed for classification based on their intra- and inter-class dependencies. These rules are then employed for the classification of a commonly used mammography dataset, and rigorous experimentation is performed to evaluate the rules' efficacy under different classification scenarios.

Mammographic image classification using histogram intersection [6], proposed using histogram intersection for mammographic image classification. It uses the bag-of-words model for image representation, which captures the texture information by collecting local patch statistics. Then, it uses normalized histogram intersection as a similarity measure with the K-nearest neighbour (KNN) classifier. Furthermore, by taking advantage of the fact that HI forms a Mercer kernel, we combine HI with support vector machines (SVM), which further improves the classification performance. They have been evaluated on a galactographic dataset and are compared with several previously used methods.

A fractal approach was proposed in [7] to model the mammographic parenchymal, ductal patterns and enhance the micro calcifications. The results proved that fractal modelling is an efficient approach for detection and classification of micro calcification in a computer aided diagnosis systems.

Classification of mammograms with benign, malignant and normal tissues using independent component was proposed by the authors in [8] with a classification accuracy of 97.3%.The face recognition methods such as AdaBost and Support vector machines are used for the analysis of digital mammograms was presented in [9].

In breast cancer diagnosis in digital mammogram uses statistical features and neural network [10], statistical features are extracted from the digital mammograms. These features are fed to neural network classifier to classify it into two classes namely normal and cancer. This study describes neural network classification technique. Experiments have been conducted on images of DDSM (Digital Database for Screening Mammography) database. The performance measures are evaluated by confusion matrix. By increasing the training samples this study reveals the improved classification accuracy.

In Features Based Mammogram Image Classification Using Weighted Feature Support Vector Machine [11], mammogram images is divided into training and test set, then the pre-processing techniques such as noise removal and background removal are applied to the input images and the Region of Interest (ROI) is identified. The statistical features and texture features are extracted from the ROI and the clinical features are obtained directly from the dataset. The extracted features of the training dataset are used to construct the weighted features and pre-computed linear kernel for training the WFSVM, from which train model file is created. Using this model file the kernel matrix of test samples are classified as benign or malignant.

In [12] proposes mammogram image segmentation using Rough K-Means (RKM) clustering algorithm. The median filter is used for pre-processing of image and it is normally used to reduce noise in an image. The 14 Haralick features are extracted from mammogram image using Gray Level Co occurrence Matrix (GLCM) for different angles. The features are clustered by K-Means, Fuzzy C-Means (FCM) and Rough K-Means algorithms to segment the region of interests for classification. The result of the segmentation algorithms compared and analyzed using Mean Square Error (MSE) and Root Means Square Error (RMSE).

In [13] focuses on classifying mammographic images based on the region of interest. The proposed classification methods include integral image computation to normalize the contrast of the image followed by gray value distribution and Histogram Equalization. Thereby, the noise is eliminated to enhance the input image for further processing. The texture features, their shape and spatial features are extracted to compute the growth of the anatomic features by using volume estimation, interest point calculation and mammogram classifier.

All these methods have the problem of overlapping and classifying false positive classes of mammographic images. We propose a multi-variant feature based classifier for mammographic images using Fuzzy c means clustering.

II. Proposed Algorithm

A. Proposed Method Architecture –

In the proposed methodology the following features namely intensity, shape, volumetric are used. The figure 1 proposed method architecture is used by preprocess the image to improve the quality of the image and then the intensity features are extracted and the shape features are extracted by identifying the connected components and its features are extracted. With the extracted features, volumetric measures are computed to classify the image.

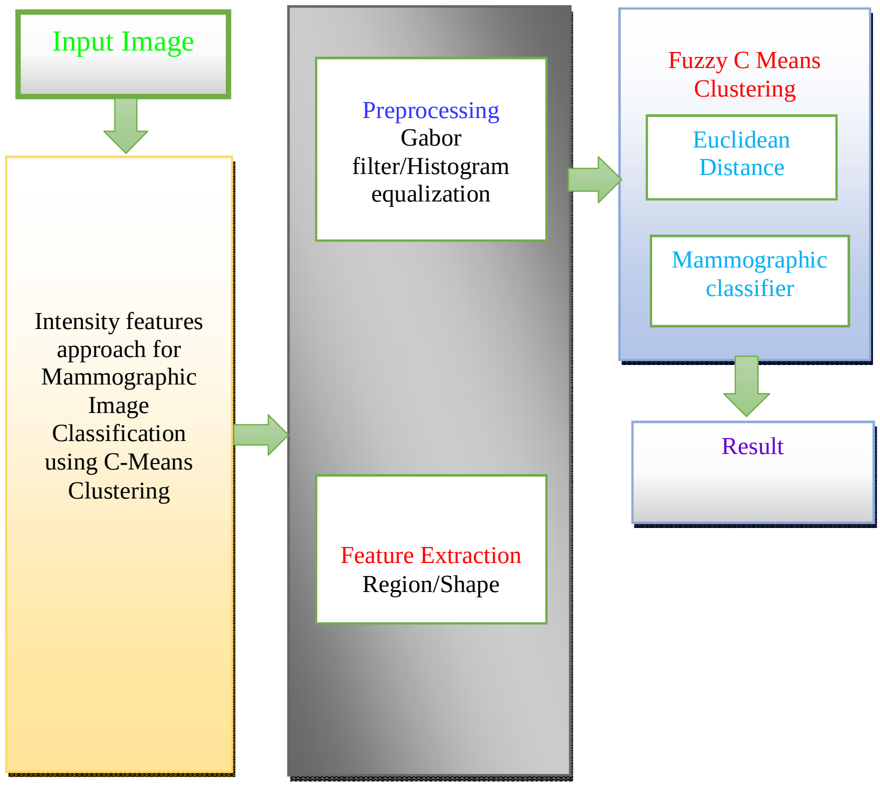


Figure1: Proposed Method Architecture

B. Pre-Processing –

The process of removing noise present in the image will be removed in this stage. The figure 2 original input image 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 which normalize the intensity values of the image in figure 3. The preprocessed 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 modeling 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.

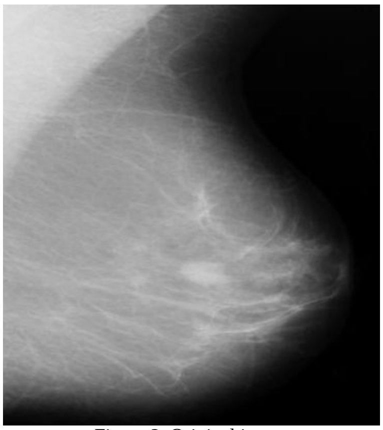


Figure 2: Original image

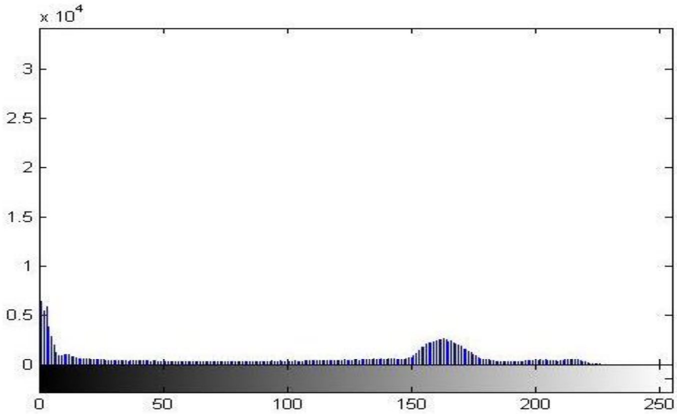


Figure3: Histogram equalization techniques image

C. Feature Extraction –

In order to provide accurate recognition, feature patterns must be extracted. Only the significant features must be encoded. In this experimentation, the method used to extract the intensity features is the Gabor filter with low and high frequencies and also with 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 we have used Gaussian filter and the filtered image is converted into binary image. From converted binary image, the regions are identified and the features. The extracted features are stored in a feature vector which will be used in further analysis.

D. Meursement Valuation –

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

E. Mammographic Classifier–

To classify the mammographic image, generated feature vectors are used and each feature vector has nine features namely intensity value, density measure, solidity, eccentricity, convex area, orientation, perimeter, major axis length, and minor axis length. With the feature vector it has number of intensity values according to the size of the pixels covered under the shape or region extracted. The mean value of intensity is computed and the density value is computed based on the pixel values present in the region. Rests of the features are left as it is to compute the Euclidean distance by the c-means clustering.

The classifier computes the Euclidean distance iteratively to compute the relative measure between the feature vectors present in the training and testing set. Based on the mean value computed with different class between feature vectors the classifier assigns a label for the submitted test image.

III. EXPERIMENT AND RESULT

The proposed methodology has been evaluated with various data sets of mammographic images. We have used the following data set for the evaluation of the proposed method.

Table 1: Shows the Details of Data set Used.

Database	Number of samples	Number of classes
DoD BCRP	750	178
MIAS	350	110
DDSM	2640	38

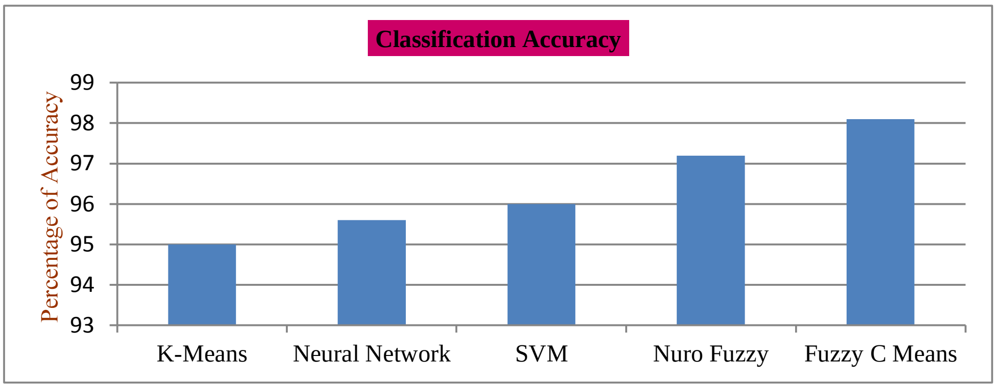
Hence, the proposed method is used from several database such as DoD BCRP which is taken from 750 number of samples. It is valued 178 number of classes in the classifications. According to Mammogram Image Analysis Society (MIAS) database is obtained of 350 number of samples and evaluated the number of classes 110.Finally, DDSM is the 38 number of classes.

The figure 4 shows the breast with cancer points and it is clear that the region marked with white color arrow represent the presence of cancer. The proposed experiment is applied to different kinds of algorithm in exiting system such as K-means, Neural Network, SVM and Nuro Fuzzy to compare with a Table2 shows the comparison of classification Accuracy.

Figure 4: shows the breast with cancer

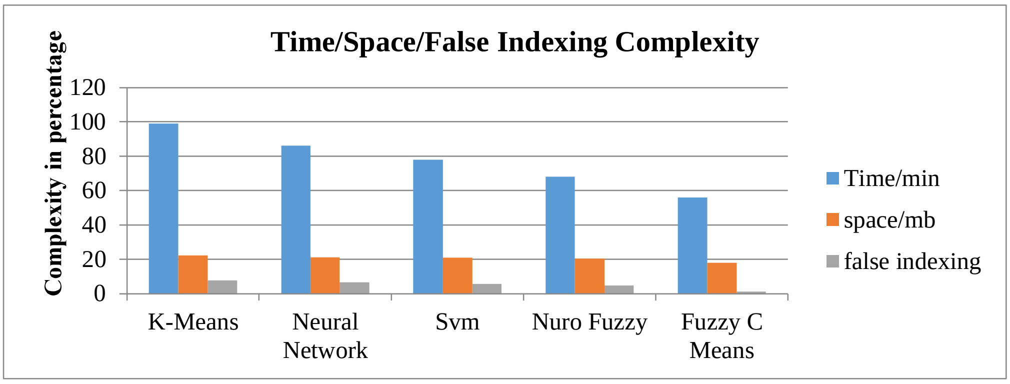
Table2: Shows the Comparison of Classification Accuracy

Algorithm	Classification Accuracy
K-Means	95
Neural Network	95.6
SVM	96
Nuro Fuzzy	97.2
Fuzzy C Means	98.1



Graph1: Shows The Classification Accuracy.

The graph1 shows the classification accuracy of different algorithms and it is clear that the proposed fuzzy c means approach has produced higher classification accuracy compare to other methods. The graph1 shows the classification accuracy for 1000 images with 3 classes. It has used 70 % of the image as training set and remaining 30 % has been used as testing set.



Graph2: Shows The Space/Time/False Indexing Complexity of Different Methods.

The graph2 shows the space and time complexity with the false indexing produced by different algorithms. It is clear that the proposed fuzzy c means approach has produced less time and space complexity with negligible false indexing which shows the more classification accuracy.

IV.CONCLUSION

This proposed a new Intensity features approach based on mammographic image classifier using fuzzy c-means which uses different features of the image like intensity values, shape and region features and density features to compute the feature vector. It has 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 98.1 % which is more than other methodologies in this era. In future, we will consider other statistical models for feature extraction in order to improve the classification rate.

REFERENCES

[1] http://www.who.int/cancer/events/breast_cancer_month/en/

[2] Erkang, C., X.Nianhua, L.Haibin, P.R.Bakic, A.D.A.Maidment, V.Megalooikonomou. Mammographic Image Classification Using Histogram Intersection. – In:IEEE Proc. ISBI'10, 2010, 197-200.

[3] Buciu, I., A. Gacsadi. Directional Features for Automatic Tumor Classification of Mammogram Images. − Biomedical Signal Processing and Control, 2011.

[4] Dua, S., S.Harpreet, H.W.Thompson. Associative Classification of Mammograms Using Weighted Rules. − Expert Syst. Appl., 2009, 9250-9259.

[5] S.Don , Duckwon Chung, A New Approach for Mammogram Image Classification Using Fractal Properties, Cybernetics and Information Technologies. Volume 12, Issue 2, Pages 69–83,2013.

[6] Erkang Cheng, Mammographic image classification using histogram intersection, IEEE Biomedical imaging, pages 197-200, 2010.

[7] HuailickJ,Fractal Modeling and Segmentation for the Enhancement of Microcalcification in Digital Mammograms. – In: IEEE Trans. Med. Imaging, 1997, 785-798.

[8] Lucio s. Diagnosis of Breast Cancer in Digital Mammograms Using Independent Component Analysis and Neural Networks. – LNCS, Vol. 3773, 2005, 460-469.

[9] ArodT, Pattern Recognition Techniques for Automatic Detection of Suspicious-Looking Anomalies in Mammograms. − Comput. Methods Prog. Biomed, 2005, 135-149.

[10] Nithya R, Santhi B (2012a). Breast cancer diagnosis in digital mammogram using statistical features and neural network. Res. J. Appl. Sci. Eng. Technol. 4(24):5480-5483

[11] S.kavitha, k.k, Thiyagarajan, Features Based Mammogram Image Classification Using Weighted Feature Support Vector Machine , Global Trends in Information Systems and Software Applications Communications in Computer and Information Science Volume 270, pp 320-329, 2012.

[12] R. Subash Chandra Boss, K. Thangavel, D. Arul Pon Daniel, Mammogram Image Segmentation using Rough Clustering, IJRET: International Journal of Research in Engineering and Technology eISSN: 2319-1163 | pISSN: 2321-7308.

[13] S.Julian Savari Antony, Improvement of Efficient Volume Estimation Mammographic Image Classification Based on Wavelet Analysis, IDEAL - VOL. - III ISSUE - I ISSN 2319 - 359X 25 SEPT.-FEB.- 2014-15.