

Possibilistic Fuzzy C Means Algorithm For Mass classificaion In Digital Mammogram

* J.S .Leena Jasmine, ** A. Safana

* Senior Grade Assistant professor, **PG scholar
Velammal Engineering College, Chennai

Abstract-

Mammography is an effective imaging modality of breast cancer abnormalities detection. Survival rate of breast cancer treatment can be increased via early detection of mammography. However detecting the mass in the early stage is a tough task for radiologist. Detection of suspicious abnormalities is a continual task. Out of thousand cases only 3 to 4 are analyzed as cancerous by a radiologist and thus abnormality may be left out. 10-30% of cancers are failed to detect by radiologist. Computer Aided Diagnosis helps the radiologists to detect abnormalities earlier than traditional procedures. Because of some negligence in capturing device, the image may be affected by noise this leads to fault diagnosis. Preprocessing can remove this unwanted noise. In this paper features such as entropy, circularity, edge detection, and correlation are extracted from the image to distinguish normal and abnormal regions of a mammogram. Classification and detection of mammogram can be done by Possibilistic Fuzzy C Means algorithm and Support Vector Machine using extracted features.

Keywords- Mammogram, Mass, Preprocessing, Feature extraction, Possibilistic Fuzzy C Means, Support Vector Machine.

I Introduction

Breast cancer is the second cause of death among women in the world [1]. The American Cancer Society estimated that 1,84,300 women will be diagnosed with breast cancer and that 44,300 women will die from it .Another study showed approximately 7,20,000 new cases will be diagnosed world-wide per year, this accounts for about 20% of all malignant tumor cases. Screening Mammography is the only method currently available for the reliable detection of early and potentially curable breast cancer.

In our human body most of the time cells split and grow in an orderly manner. In some cases growth of cells increases without control. This will lead to formation of mass or lump known as tumor. Tumors can also be benign or malignant. Benign tumors are not considered as cancerous. Benign tumor cells do not attack nearby tissue or spread to other parts of the body. When these tumors are removed, most of the time they do not come back. Breast cancer is any form of malignant tumor which develops from breast cells [8].The abnormalities in mammograms can be roughly divided into two types: micro calcifications or masses [4].

Several databases have commonly been used as test beds for the performance of the proposed segmentation algorithms. A large number of images are necessary to test a Computer Aided Diagnosis system and to compare processing results with others for performance evaluations. In order to overcome the difficulty in accessing hospitals and

clinics confidential files, there is a need for a public database. MIAS (Mammographic Image Analysis Society Digital Mammogram Database) is well known and broadly used mammographic database. MIAS database is in pgm format with 8 bits images.

II Literature Review

Serifovic Trbalic et al. [3] developed computer-aided diagnosis system for classification of the abnormalities in the mammograms as malignant or benign still remains a challenge in the digital mammography. The input Regions of Interest (ROIs) are obtained using an efficient Otsu's N thresholding and further subjected to a number of preprocessing stages. After preprocessing stage, from the ROIs, a group of 32 Zernike moments with different orders and iterations have been extracted. These moments have been applied to the neural network classifier.

V.Thavavel et al. [7] proposed a method for early detection of cancerous regions in mammogram images using mathematical morphological and rough set based approach in detection and classification of cancerous masses in MRI mammogram images.MRI Mammogram images is enhanced and the artifacts are removed using the fuzzification technique. The ROI is extracted using Graph Cut method and the four mathematical morphological features are calculated for the segmented contour. ID3 algorithm is applied to extract the features which play a vital role in classification of masses in mammogram into Normal, Benign and Malignant.

G. Balakrishnan et al. [6] presented a paper for mass classification of mammogram based on Discrete Wavelet Transform (DWT) and Stochastic Neighbor Embedding (SNE) Technique. The mass classification in digital mammogram is achieved by decomposing the mammogram image by using DWT at various levels. Then the high dimensional wavelet coefficients are reduced by using SNE. The reduced wavelet coefficients are used as features for the corresponding mammogram image to classify the given mammogram into normal, benign and malignant mass.

L.Megalan Leo et al. [12] states that noise presence in the image and the nature of the varying intensity makes segmentation process a crucial one. Some of the approaches used over this problem are region growing, watershed segmentation, global thresholding, and active contour models. Active contour methods are widely used for the image segmentation because of the advantageous properties such as topology adaptability, and robustness to initialization. In this paper they present and demonstrate the effectiveness of some active contour models on the mammogram images where edges are weak. Finally they plan to suggest a method which is efficiently segment the mammogram images based on the result of the contour models.

III Methodology

A. Possibilistic Fuzzy C Means

The clustering algorithm used in this paper is Possibilistic Fuzzy C Means. PFCM is a good clustering algorithm to perform classification tests because it possesses capabilities to give more importance to typicalities or membership values [13]. In order to avoid the constraint corresponding to the sum of all typicality values of all data to a cluster must be equal to one cause problems particularly for a big data set Pal et al. [15] propose a new algorithm called Possibilistic Fuzzy c means algorithm. It produces memberships and possibilities simultaneously, with the usual point prototypes or cluster centers for each cluster. The first step of PFCM is an initialization step in which randomly set the membership/typicality degrees and centroids. After that, the membership/typicality degrees and centroids are updated iteratively, when the changes in the centroids are insignificant then the iterative process will stop. This means that the regions become stable. The final step is assigning all pixels in the input image to the object having the highest degree of typicality and membership. Hence, and as a result of this clustering process, we obtain a set of objects each modeled with a subset of pixels (or region) of the original image. PFCM, like any partitioning-based method, requires specification of the number of objects in the image. This number

could be computed using a validity measure as the information entropy, variance, and so on.

The objective function is defined by

$$J_{pfc m} = \sum_{i=1}^c \gamma_i \sum_{k=1}^n (a\mu_{ik}^m + bt_{ik}^\eta) \|z_k - v_i\|^2 + \sum_{k=1}^n \sum_{i=1}^c (1 - t_{ik})^\eta \quad (1)$$

Subject to the constraints $\sum_{i=1}^c \mu_{ik} = 1$ for all k and $0 \leq \mu_{ik}, t_{ik} \leq 1$.

Here $a > 0, b > 0, m > 0$, and $\eta > 0$. U is the partition matrix. T is the typicality matrix. V is a vector of cluster centers, X is a set of all data points, z represents a data point, n is the number of data points and c is the number of cluster centers which are described by s coordinates

Algorithm:

Step 1: Initialize prototype

$$v_i = \sum_{k=1}^N (a\mu_{ik}^m + bt_{ik}^\eta) z_k / \sum_{k=1}^N (a\mu_{ik}^m + bt_{ik}^\eta), \quad 1 \leq i \leq C \quad (2)$$

Step 2: For each cluster compute penalty parameter γ_i

$$\gamma_i = K \frac{\sum_{k=1}^N \mu_{ik}^m \|z_k - v_i\|^2}{\sum_{k=1}^N \mu_{ik}^m} \quad (3)$$

Step 3: For each prototype calculate the distance

$$D_{ikA}^2 = (z_k - v_i)^T A_i (z_k - v_i), \quad 1 \leq i \leq C, 1 \leq k \leq N \quad (4)$$

Step 4: Calculate membership and typicality values

$$\mu_{ik} = \left(\sum_{j=1}^c \left(\frac{D_{ikA_i}}{D_{jkA_j}} \right)^{2/(m-1)} \right)^{-1}, \quad 1 \leq i \leq C, 1 \leq k \leq N \quad (5)$$

$$t_{ik} = \frac{1}{1 + \left(\frac{b}{\gamma_i} D_{ikA_i}^2 \right)^{1/(\eta-1)}}, \quad 1 \leq i \leq C, 1 \leq k \leq N \quad (6)$$

Step 5: Update the values of prototypes.

B. Support Vector Machine

SVM is a useful technique for data classification. In this paper we form a training set using extracted feature values and determine whether the mammogram is in benign, normal or malignant. Support vector machine can be used only when the data has exactly two classes. A classification task usually involves with training and testing data which consist of some data instance. Each instance in the training set contains one target values and more than a few attributes. The aim of SVM is to make a model which predicts target value of data instances in the testing set which are given only the attributes. A step in SVM classification

involves identification as which are intimately connected to the identified classes. This is known as feature selection or feature extraction. Feature selection and SVM classification together have a use

An SVM classifies data by finding the best hyperplane that separates all data points of one class from those of the other class. The finest hyperplane for an SVM means the one with the largest margin between the two classes. Margin means the maximal width parallel to the hyperplane that has no interior data points. The support vectors are the data points that are closest to the separating hyperplane.

IV Proposed System

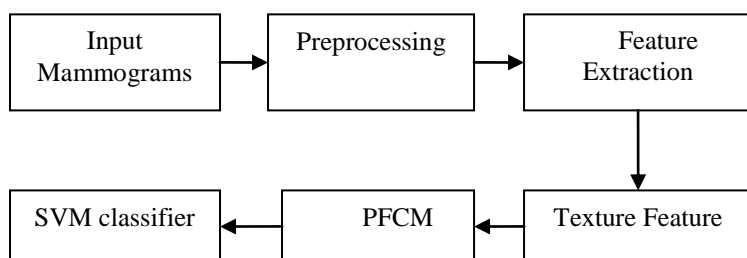


Figure 1: Procedure for mass classification

A. Preprocessing

The preprocessing algorithm, techniques and operators are used to perform initial processing that makes the primary data reduction and analysis task easier. It is an important issue in low-level image processing. They include operations related to:

- Extracting regions of interest.
- Performing basic algebraic function on image.
- Enhancing specific image features.
- Reducing resolution and brightness of data.

Adaptive Median filtering

In this paper we use adaptive median filter to remove the unwanted noise. It performs spatial processing to determine which pixel in an image has been affected by impulse noise. It classifies pixels as noise by comparing each pixel in the image to its surrounding neighbors. A pixel that is different from majority of its neighbors, as well as being not structurally aligned with those pixels to which it is related, is identified as impulse noise. These noise pixels are then replaced by the median pixel value of pixels in the neighborhood that have passed the noise labeling test.

The Median filter takes an area of image, observes all pixel values in that area and puts it into the array called element array. Then, the element array is sorted and the median value of the element array is found out. We have achieved this by sorting

even when prediction of unknown samples is not required. They can be used to recognize key sets which are involved in whatever Processes distinguish the classes.

the element array in the ascending order using bubble sort and returning the middle elements of the sorted array as the median value. Then the output image is the set of all the median values of the element arrays obtained for all the pixels.

Image Enhancement

Improves the quality of image. Images are enhanced before the screening process. Enhancement procedures can amplify certain features or content in an image with the goal of improving screening effectiveness. Here spatial domain method is used, technique based on direct manipulation of pixels in an image. Image enhancement can also be used to diminish certain image features or content, such as noise reduction and/or background removal. Mammogram images may be enhanced before being processed by other screening algorithms or the enhanced images may be prepared for direct viewing by radiologists. To process an image so that the result is more suitable than the original image for a specific application. In order to make images more visually appealing and highlighting interesting detail in images this process can be done.

B. Feature Extraction

For classification, extraction of features is important. If the features extracted are carefully chosen it is expected that the features set will extract the relevant information from the input data in order to perform the desired task using this reduced representation instead of the full size input. The idea is to calculate the co-occurrence matrix for small regions of the image and then use this matrix to find statistic values, for instance, Correlation, Eccentricity, Inverse and Entropy. Features can be a set of real numbers which characterize a normal mammogram. Some simple statistical features could be the mean and standard deviation of pixel intensities of a region. However, feature extraction process should be problem-dependent.

C. SVM Algorithm

In our proposed method we use Support Vector Machine for classification of mammogram images. The steps for doing SVM are as follows

Step 1: Load the sample data

load dataname

Step 2: Create data, a two-column matrix

data = [mias(:,1), mias(:,2)];

step 3: From the species vector, create a new column vector, groups, to classify data into two groups: data and non-data.

```
groups = ismember(dataset,'data');
```

step 4: Randomly select training and test sets.

```
[train, test] = crossvalind('holdOut',groups);
```

```
cp = classperf(groups);
```

Step 5: Use a linear support vector machine classifier

```
svmStruct = svmtrain(data(train,:),groups(train)
'showplot',true);
```

step 6: Use the svmclassify function to classify the test set.

```
classes =
```

```
svmclassify(svmStruct,data(test,:), 'showplot',true
);
```

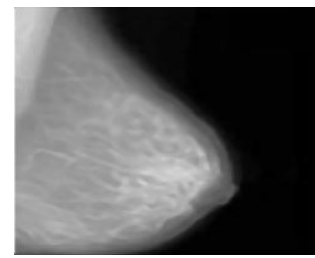
step 7: Evaluate the performance of the classifier.

```
classperf(cp,classes,test);
```

```
cp.CorrectRate
```

IV Experimental Result

Read the user selected input image and convert into binary image Fig .a. Fig .b shows the perimeter of image. Then resize them into pixel range of 255X255 fig .c. By using adaptive median filtering technique removal of noise in an image and enhancing can be shown by fig .d and fig .e. Find Feature Extraction that may help to find out the exact affected region and values, such as Area, Perimeter, circularity, Eccentricity, Elongation, Rectangularity, Volume Tissue, Volume whole, Glcm, Energy, Correlation, Sum Of Variances, Mean, Standard Deviation, Entropy Of The Image, Smoothing of the Image, Centroid.



(e)

Figure 2 : (a)Input image (b) perimeter image (c) Resized Image (d) Filtered Image (e) Enhanced Image

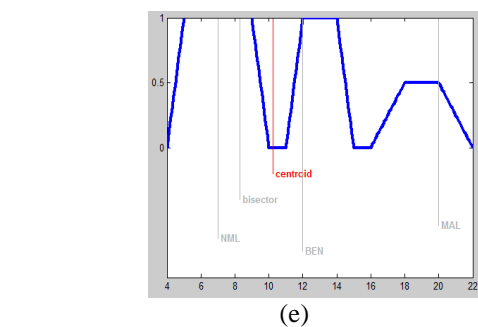
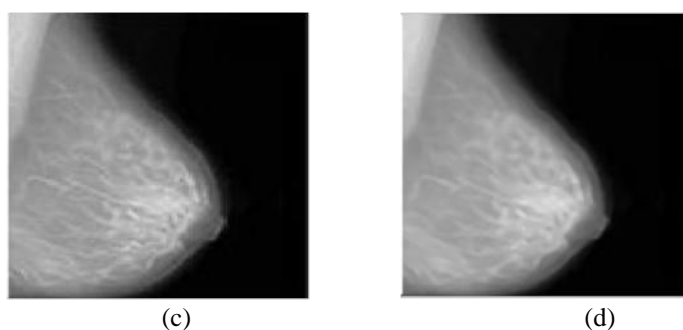
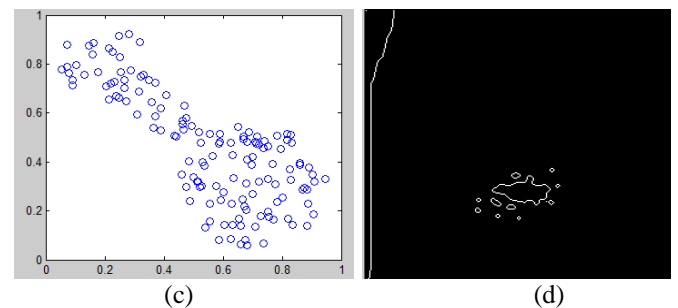
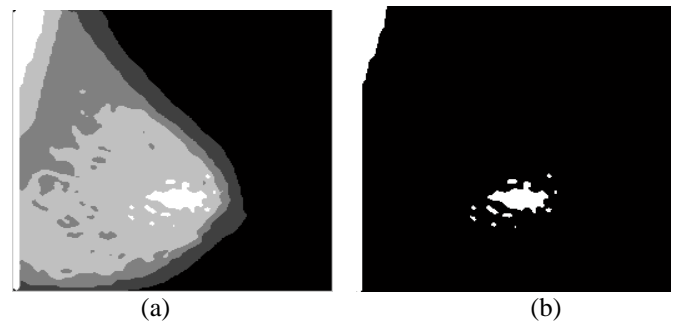


Figure 3: (a) Clustered Image (b)Output Image (c) Input Feature Vector (d) Clustered Feature Vector (e) Location of Centroid

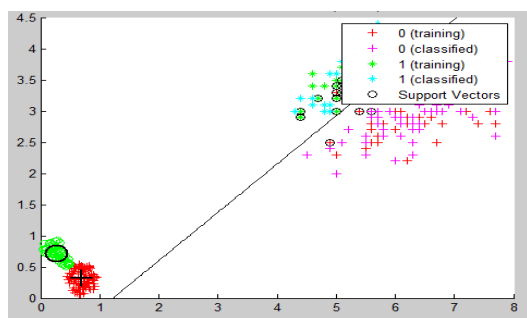


Figure 4: Support Vector Machine

Fig.3.represent Particular affected area in mammogram image . Based on that features find FCM classification, PCM Classification, PFCM Classification. Then find cluster PFCM Feature Vector Classification fig d. Accurate location of centroid in the affected area is represented by fig e. By using the support vector machine the clustered feature vector of image can be shown in fig.4.

S.NO	PARAMETERS	EXTRACTED VALUES
1	Area	3.3617e+004
2	Perimeter	31.9341
3	Circularity	0.0303
4	Rectangularity	0.7165
5	Volume Tissue	0.001122645136486
6	Volume Whole	1.537870049980777e-005
7	GLCM	61802 82 333 2533
8	Energy	0.912034101050536
9	Correlation	2.090552367392599
10	Sum of variance	0.673894163964799
11	Mean	0.996796356337811
12	Standard deviation	0.20594437
13	Entropy of the image	0.19945029
14	Smoothing of the image	12.00000000

Table 1 : Extracted Features

V Conclusion

In this project, a new method of classification and detection of mammogram is presented. The result shows that the proposed method improves the accuracy. Although proposed system ignores the noise sensitivity deficiency and eliminates the row of sum constraint. Until now, there has been no paper that has done classification of mammogram by using the algorithm applied here. The experimental result shows that the proposed technique is very effective in classifying the mammogram image better than existing method.

References

- [1] Khalid El Fahssi, Abdelali Elmoufidi, Abdenbi Abenaou, Said Jai-Andaloussi, Abderrahim Sekkaki “Feature extraction of the lesion in [2]Bommeswari Barathi, SivaKumar.R, Karnan.M “Computer Aided Detection Algorithm for Digital Mammogram Images – A Survey”, IJCTT, vol 8, No.3, 2014.
- [3] A.Serifovic Trbalic, A.Trbalic, D.Demirovic, N.Prljaca, P.C.cattin “Classification of benign and malignant masses in breast mammograms”, MIPRO, 2014.
- [4] Shen-Chuan Tai, Zih-Siou Chen, Wei-Ting Tsai “An Automatic Mass Detection System in Mammograms Based on complex Texture Features”, IEEE, VOL. 18, NO. 2, MARCH 2014.
- [5] M. H. Fazel Zarandi M. Rostam Niakan Kalhori M. F. Jahromi “Possibilistic C-Means Clustering Using Fuzzy Relations”, IEEE, 2013.
- [6] S Mohan Kumar, G. Balakrishnan, “Multi Resolution Analysis for Mass Classification in Digital Mammogram using Stochastic Neighbor Embedding”, International conference on Communication and Signal Processing 2013.
- [7] S.Pitchumani Angayarkanni, DR.Nadira Banu Kamal, and DR.V.Thavavel “Mammogram MRI Image Segmentation and Classification, JATIT 2013.
- [8] Vicky Ambule, Minal Ghute, Kanchan Kamble, Shilpa Katre “Adaptive Median Filter for Image Enhancement”, IJESIT, Volume 2, Issue 1, January 2013.
- [9] Karthikeyan Ganesan, U. Rajendra Acharya, Chua Kuang Chua, Lim Choo Min, K. Thomas Abraham, and Kwan-Hoong Ng “Computer-Aided Breast Cancer Detection Using Mammograms: A Review”, IEEE Reviews in Biomedical Engineering, Vol. 6, 2013.
- [10] Hamed Shamsi, Hadi Seyedarabi “A Modified Fuzzy C-Means Clustering with Spatial Information for Image Segmentation”, IACSIT, Vol. 4, No. 5, 2012.
- [11] R.Suganya, R.Shanthi “Fuzzy C- Means Algorithm- A Review”, International Journal of Scientific and Research Publications, Volume 2, Issue 11, 2012.
- [12] Mr. L.Megalan Leo, Mr. A.Ranjith, Mr. R.Thandaiah Prabu “Analysis of Mammogram Images using Active Contour Segmentation process and Level Set Method, IJTAE 2012.

- [13] Neelam Kumari, Bhawna Sharma, Dr. Deepti Gaur “ Implementation of Possibilistic Fuzzy C-Means Clustering Algorithm in Matlab”, International Journal of Scientific & Engineering Research, Volume 3, Issue 11, November-2012.
- [14] P. Hari Krishnan, Dr. P. Ramamoorthy “An Efficient Modified Fuzzy Possibilistic C-Means Algorithm for MRI Brain Image Segmentation”, IJERA, Vol. 2, Issue 2, Mar-Apr 2012.
- [15] Nikhil R. Pal, Kuhu Pal, James M. Keller, and James C. Bezdek,” A Possibilistic Fuzzy c-Means Clustering Algorithm ”, IEEE Trans. on Fuzzy Systems, vol. 13, no. 4, pp. 517-530, Aug. 2005.