

International Journal of Computational Intelligence and Informatics, Vol. 3: No. 3, October - December 2013 Mammogram Image Feature Extraction using Pulse-Coupled Neural Network

R Subash Chandra Boss

Department of Computer Science Periyar University, Salem-636011 Tamil Nadu, India rmsubash_18@yahoo.co.in

C Velayutham

Department of Computer Science, Aditanar College of Arts and Science, Virapandianpatnam- 628216 Tamil Nadu, India. cvelayutham22@yahoo.com

K Thangavel

Department of Computer Science Periyar University, Salem-636011 Tamil Nadu, India drktvelu@yahoo.com

D Arul Pon Daniel

Department of Computer Science Periyar University, Salem-636011 Tamil Nadu, India apdaniel86@yahoo.com

Abstract-A typical mammogram image processing system generally consists of mammogram image acquisition, pre-processing, segmentation, feature extraction, feature selection and classification. The texture description methods such as GLCM, GLDM, SRDM, and GLRLM are widely used to extract features in mammogram images for analysis and identification of micro calcification. The Pulse-Coupled Neural Networks (PCNN) is found a very good feature extraction model widely used to extract features in some of the images. The PCNN features are extracted from the mammogram images and analyze classification performance along with GLCM, GLDM, SRDM, and GLRLM features, extracted from the same mammogram images. These processes are executed and analyses the features. The performance of the proposed PCNN Feature extraction Method is examined and the experimental results are illustrated.

Keywords- Mammogram image, PCNN, GLCM, GLDM, SRDM and GLRLM Feature Extraction, Classification

I. INTRODUCTION

The high incidence of breast cancer in women, especially in developed countries, has increased significantly in the last one decade. Though much less common, breast cancer also occurs in men [1, 2]. The etiologies of this disease are not clear and neither are the reasons for the increased number of cases. Currently there are no methods to prevent breast cancer, which is why early detection represents a very important factor in cancer treatment and allows reaching a high survival rate. Mammography is considered as the most reliable method in early detection of breast cancer. Due to the high volume of mammograms to be read by physicians, the accuracy rate tends to decrease, and automatic reading of digital mammograms becomes highly desirable. It has been proven that double reading of mammograms (consecutive reading by two physicians or radiologists) increased the accuracy, but it involves high costs. That is why the computer aided diagnosis systems are necessary to assist the medical staff to achieve high efficiency and effectiveness.



Figure 1. Image categorization process

The texture description methods such as GLCM, GLDM, SRDM, and GLRLM are widely used to extract features in mammogram images for analysis and identification of micro calcification [3, 4]. The Pulse-Coupled Neural Networks (PCNN) method can be found a very good feature extraction model widely used in the area of image processing. The PCNN features are extracted from the mammogram images and analyses of classification performance along with GLCM, GLDM, SRDM, and GLRLM features extracted from the same mammogram images are performed. A typical mammogram image processing system generally consists of mammogram image acquisition, pre-processing, segmentation, feature extraction, feature selection and classification. These processes are executed and analyses the features in this paper. Fig. 1 shows Image categorization process.

The rest of the paper is organized as follows Section 2, presents the mammogram image acquisition. Section 3 describes the image pre-processing and image segmentation process. Section 4 describes feature extraction methods Section 5, describes the proposed PCNN model, Section 6 describes feature selection. Section 7 presents FCM method for classification of images. Section 8 presents the experimental results and comparative analysis. Conclusions are presented in section 9.

II. MAMMOGRAM IMAGE ACQUISITION

The mammogram images used for experimental analysis are taken from the Mammographic Image Analysis Society (MIAS) [5]. Its corpus consists of 322 images, which belong to three big categories: normal, benign and malign. There are 208 normal images, 63 benign and 51 malign, which are considered abnormal. In addition, the abnormal cases are further divided into six categories: microcalcification, circumscribed masses, speculated masses, ill-defined masses, architectural distortion and asymmetry. All the images also include the locations of any abnormalities that may be present. The existing data in the collection consists of the location of the abnormality (like the centre of a circle surrounding the tumour), its radius, breast position (left or right), type of breast tissues (fatty, fatty glandular and dense) and tumour type if exists (benign or malign).

III. PRE-PROCESSING AND IMAGE SEGMENTATION

Mammogram images are difficult to interpret, and a preprocessing of the images is necessary to improve the quality of the images and make the feature extraction more reliable. Pre-processing is always a necessity whenever the data to be mined in noisy, inconsistent or incomplete and pre-processing significantly improves the effectiveness of the data mining techniques [6]. This section introduces the pre-processing techniques applied to the images before the feature extraction. In the digitization process, noise could be introduced that needs to be reduced by applying some image processing techniques. In addition, at the time that the mammograms were taken, the conditions of illumination are generally different. The cropping operation was employed in order to cut the black parts of the image as well as the existing artefacts such as written labels etc. For most of the images in this dataset, almost 50% of the whole image comprised of a black background with significant noise. Cropping removed the unwanted parts of the image and cutting horizontally and vertically the image those parts that had the mean less than a certain threshold.

Image enhancement helps in qualitative improvement of the image with respect to a specific application [7]. In order to diminish the effect of over brightness or over darkness in the images and accentuate the image features, we applied a widely used technique in image processing to improve visual appearance of images known as Histogram Equalization. Histogram equalization increases the contrast range in an image by increasing the dynamic range of grey levels (or colours) [7]. This improves the distinction of features in the image. The method proceeds by widening the peaks in the image histogram and compressing the valleys. This process equalizes the illumination of the image and accentuates the features to be extracted. (i.e) how the different illumination conditions at the scanning are reduced. The suspicious region or microcalcifications is segmented using optimization algorithm for mammogram images [8]. Fig. 2 shows the example of Image pre-processing and segmentation process.



Figure 2. Example of Image pre-processing

IV. FEATURE EXTRACTION

The texture of mammogram images refers to the appearance, structure and arrangement of the parts of an object within the image. A feature value is a real number, which encodes some discriminatory information about properties of an object. Texture is one of the important characteristics used in identifying an object [9]. The texture coarseness or fineness of an image can be interpreted as the distribution of the elements in the matrix such as GLCM, GLDM, SRDM, and GLRLM.

The texture analysis matrix itself does not directly provide a single feature that may be used for texture discrimination. Instead, the matrix can be used as a representation scheme for the texture image and the features are computed from the texture discrimination matrix.

A. Gray Level Co-occurrence Matrix (GLCM)

Generally, the problem of texture discrimination based on statistical approach consists of the analysis of a set of co-occurrence matrices [10, 11]. In this matrix, the indices of rows and columns represent the given range of the image gray levels, and the value P (i, j) stored at the position (i, j) is the frequency that gray levels i and j occur with, at a given distance and at a given direction. For instance, suppose we have the image represented by its gray level matrix. Regarding the angle at 0, 45, 90, and 135 degrees . Fig.3a shows its direction and the distance one, we will have the gray level co-occurrence matrices as shown in the Table-I.

Illustration for Gray Level Co-occurrence Matrix



The matrix is dividing each element by normalization factors renders a matrix whose sum is one.

| Deg | (i) | 0 | 1 | 2 | 3 | |
|-----------------|------------|---|----|---|---|--|
| Des | (i) | Ŭ | - | - | e | |
| 0 | 0 | 2 | 9 | 3 | 1 | |
| U | 1 | 9 | 6 | 2 | 7 | |
| | 2 | 3 | 2 | 4 | 0 | |
| | 3 | 1 | 7 | 0 | 4 | |
| 45 [°] | 0 | 4 | 2 | 1 | 5 | |
| 4.5 | 1 | 2 | 14 | 1 | 3 | |
| | 2 | 1 | 1 | 4 | 1 | |
| | 3 | 5 | 3 | 1 | 2 | |
| 90° | 0 | 4 | 7 | 2 | 0 | |
| | 1 | 7 | 10 | 2 | 8 | |
| | 2 | 2 | 2 | 2 | 2 | |
| | 3 | 0 | 8 | 2 | 2 | |
| 135 | 0 | 4 | 4 | 2 | 2 | |
| 100 | 1 | 4 | 8 | 4 | 5 | |
| | 2 | 2 | 4 | 0 | 1 | |
| | 3 | 2 | 5 | 1 | 2 | |

B. Gray Level Difference Matrix (GLDM)

T . C (

The GLDM is based on the occurrence of two pixels which have a given absolute difference in gray level and which are separated by a specific displacement δ . For any given displacement vector $\delta = (\Delta x, \Delta y)$.

Let
$$S(x, y) = |S(x, y) - S(x + \Delta x, y + \Delta y)$$
 (1)

and the estimated probability-density function is defined by

. .

$$(i \mid \delta) = \operatorname{Prob}\left(S_{o}\left(x, y\right) = 1\right) \tag{2}$$

Table II shows the GLDM of the sample image matrix Fig. 3.

| i δ | 0° | 45 | ° 90 | [°] 135 | • |
|--------|----------|---------|---------|------------------|---|
| 1 | 16 22 | 12 4 | 9 11 | 7 9 | |
| 3 4 | 20 2 | 4 5 | 10 0 | 7 2 | |

C. Surrounding Region Dependency Matrix (SRDM)

The SRDM is based on a second-order histogram in two surrounding regions. The mammogram image is transformed into a surrounding region-dependency matrix and the features are extracted for this matrix. Let us consider two rectangular windows centered on a current pixel (x, y). R_1 and R_2 are the outermost and outer surrounding region of size 7x7 and 5x5, respectively. The number of pixels greater than the selected threshold value (q) is counted in each region. Let us assume m and n to be the total number of pixels from the outermost region (R_1) and the outer region (R_2) . The element in the corresponding surrounding region dependency matrix M (m, n) is incremented by 1. This procedure is repeated for all the image pixels and the matrix gets updated.

The SRDM matrix is generated for certain threshold value. The SRDM matrix has the dimension of $m \times n$, where m is the total number of pixels in the R_1 region and n is the total number of pixels in the R_2 region. For example, if the threshold value is 1, R_1 contains 16 pixels and region R_2 contains 10 pixels having greater intensity values than the threshold value. So, the value of $(16, 10)^{th}$ element in the SRDM matrix is incremented by one as M (16,10) = M (16,10) + 1. Fig. 4 shows a typical SRDM matrix for the sample image matrix 1 in Fig.3b.



| | | | | | (; | a) | | | | | |
|---|---|---|---|---|----|----|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | | | | (1 | 5) | | | | | |

Figure 4 (a) and (b): SRDM of sample image matrix

D. Gray Level Run Length Matrix

A run length is represented by (i, j, θ), where i, j, and θ are the gray level, run length, and direction, respectively. Run lengths carry texture information on both direction and coarseness. For simplicity, 0°, 45°, 90°, and 135° are the most widely used directional parameters. Fig. 3a shows these directions. Gray level run length features are a form of gray level statistical features. A gray level image can be decomposed into gray level runs, where each run is a series of consecutive pixels of the same intensity in some predefined orientation. Run lengths are well defined in raster images at orientations in increments of 45 degrees starting at 0 degrees. The run is represented by the pixel intensity and the length of the run in pixels. The process of decomposing an image in this way is known as run length encoding. A run length encoding of a gray level image is often significantly smaller than the source image, so the process has been used as a simple image compression scheme. A gray level run length matrix is computed from the run length encoding of the image. Let m be the number of gray levels in the image and n be the length of the longest run. The gray level run length matrix is an m x n matrix where the (i, j)th element P_{ij} is the number of runs with intensity i and length j.

1) Illustration for Gray Level Run Length Matrix

A sample image matrix in Fig. 3a is used to illustrate the run length measures. Its run length measures at the four directions are shown in Table III, where i and j stand for gray level and run length, respectively.

| | | • | 1 0 | 2 | |
|-----------------|-------|----|-----|---|---|
| | | U | 1 2 | 3 | |
| | (J) \ | | | | |
| റ് | 1 | 6 | 1 | 0 | 0 |
| v | 2 | 10 | 3 | 0 | 0 |
| | 3 | 1 | 2 | 0 | 0 |
| | 4 | 3 | 2 | 0 | 0 |
| 45 [°] | 1 | 4 | 2 | 0 | 0 |
| 43 | 2 | 6 | 1 | 0 | 2 |
| | 3 | 1 | 2 | 0 | 0 |
| | 4 | 5 | 1 | 0 | 0 |
| 000 | 1 | 4 | 2 | 0 | 0 |
| 90 | 2 | 7 | 3 | 1 | 0 |
| | 3 | 3 | 1 | 0 | 0 |
| | 4 | 5 | 1 | 0 | 0 |
| 125 | 1 | 4 | 2 | 0 | 0 |
| 155 | 2 | 9 | 2 | 1 | 0 |
| | 3 | 5 | 0 | 0 | 0 |
| | 4 | 5 | 1 | 0 | 0 |

 TABLE III.
 GLRLM OF SAMPLE IMAGE MATRIX

E. The Haralick Features

The features based on the distribution matrices should therefore capture some characteristics of textures such as homogeneity, coarseness, periodicity and others. Haralick et al. have suggested 14 texture features [12]:

| i) Angular Second Moment (ASM) | $\sum_{i} \sum_{j} (p(i,j))^2$ | vii) Sum Variance | $\sum_{i=2}^{2N_g} (i - f_g)^2 p_{x+y}(i)$ |
|-----------------------------------|--|--|--|
| ii) Contrast | $\sum_{n=0}^{N_g-1} n^2 (\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i,j))$ | viii) Sum Entropy | $-\sum_{i=2}^{2N_g} p_{x+y}(i) \log(p_{x+y}(i))$ |
| iii) Correlation | $\frac{\sum_{i=j} \sum_{j=1}^{j} (i,j) p(i,j) - \mu_x \mu_y}{\sigma \sigma}$ | ix) Entropy | $-\sum_{i}\sum_{j}p(i,j)\log(p(i,j))$ |
| | where $\mu_x = \sum_{i=1}^{n-1} i p_x(i)$, | x) Difference Variance | $Variance_of(p_{x-y})$ |
| | $\mu = \sum_{i=0}^{n-1} i p_i(i)$ | | $-\sum_{i=0}^{s} p_{x-y}(i) \log(p_{x-y}(i))$ |
| | $\sigma_x^2 = \sum_{j=0}^{n-1} (i - \mu_x)^2 p_x(i)'$ | xii) Information measures of Correlation-I | $\frac{HXY - HXY1}{\max(HX, HY)}$ |
| | $\sigma_{y}^{2} = \sum_{j=0}^{n-1} (j - \mu_{y})^{2} p_{y}(j)$ | | where $HXY = -\sum \sum p(i, j) \log(p(i, j))$ |
| iv) Sum of squares: Variance | $-\sum_{i}\sum_{j}(i-v)^{2} p(i,j)$ | | |
| v) Inverse Difference Moment | $-\sum_{i}\sum_{j}\frac{1}{1+(i-j)^{2}}p(i,j)$ | | $HXY_1 = -\sum_i \sum_j p(i, j) \log(p_x(i)p_y(j))$ |
| vi) Sum Average | $\sum_{i=2}^{2N_g} ip_{x+y}(i)$ | | $HXY_2 = -\sum_i \sum_j p_x(i) p_y(j) \log(p_x(i) p_y(j))$ and HX and HY are entropies of p_x and p_y |

| xiii) Information measures of Correlation-II | $(1 - \exp\left[-2.0(HXY2 - HXY)\right])^{\frac{1}{2}}$ where $HXY = -\sum_{i} \sum_{j} p(i, j) \log(p(i, j))$ |
|--|--|
| | $HXY_{1} = -\sum_{i} \sum_{j} p(i, j) \log(p_{x}(i)p_{y}(j))$ |
| | $HXY_{2} = -\sum_{i} \sum_{j} p_{x}(i) p_{y}(j) \log(p_{x}(i)p_{y}(j))$ |
| | and HX and HY are entropies of p_x and p_y |

| Correlation coefficient(Second largest eigenvalue of $Q^{1/2}$ where $Q(i, j) = \sum_{k} \frac{1}{p_{x}(i)}$ | $\frac{k)p(j,k)}{(i)p_{y}(k)}$ |
|---|--------------------------------|
|---|--------------------------------|

All these extracted features are computed over smaller windows of the segmented image.

F. Gray Level Run Length Features

GLRL Features is computed based on run length decompositions in all of the four orientations 0, 45, 90, and 135 degrees. (The decompositions at 180, 225, 270 and 315 yield identical gray level run length matrices to those at 0, 45, 90, and 135 degrees respectively). GLRL features capture both structural and statistical information from the texture [13], and therefore provide some improvement over purely statistical features. Unfortunately, only relationships between pixels of the same intensity are captured, and all other relationships are ignored.

| i) Short Run Emphasis | $\frac{\sum_{i}\sum_{j}p(i,j)}{j^2}$ |
|-----------------------------------|---|
| ii) Long Run Emphasis | $\frac{\sum_{i}\sum_{j}j^{2}p(i,j)}{\sum_{i}\sum_{j}p(i,j)}$ |
| iii) Gray Level Non-uniformity | $\frac{\sum\limits_{j} (\sum\limits_{i} p(i,j))^2}{\sum\limits_{i} \sum\limits_{j} p(i,j)}$ |
| iv) Run Length Non-uniformity | $\frac{\sum\limits_{i} (\sum\limits_{j} p(i, j))^2}{\sum\limits_{i} \sum\limits_{j} p(i, j)}$ |
| v) Run Percentage | $\frac{\sum_{i} \sum_{j} p(i, j)}{n.}$ where n=number of pixels in the image |
| vi) Low Gray Level Run Emphasis | $\frac{\sum_{i} \sum_{j} \frac{p(i, j)}{i^2}}{\sum_{i} \sum_{j} p(i, j)}$ |
| vii) High Gray Level Run Emphasis | $\frac{\sum_{i}\sum_{j}i^{2}p(i,j)}{\sum_{i}\sum_{j}p(i,j)}$ |

V. PCNN MODEL

The Pulse-Coupled Neural Network (PCNN) was derived from studies on the cat's eye. It was Eckhorn who first made a model of the cat's visual cortex [7]. There are several compartments in the Eckhorn neuron. It has two input compartments, linking and feeding. The feeding compartment (F) receives both an external and a local stimulus, whereas the linking compartment (L) only receives a local stimulus. The feeding and linking are combined to form the membrane voltage U. This is then compared to a local threshold E. the structure of PCNN is given in Fig. 5a.



Figure 5a. PCNN structure

 $F_{ij}(n) = e^{-\alpha_F} F_{ij}(n-1) + V_F \sum_{kl} M_{ijkl} Y_{kl}(n-1) + S_{ij}$ $L_{ij}(n) = e^{-\alpha_L} L_{ij}(n-1) + V_L \sum_{kl} W_{ijkl} Y_{kl}(n-1)$ $U_{ij}(n) = F_{ij}(n) (1 + \beta L_{ij}(n))$ $Y_{ij}(n) = \begin{cases} 1 & , U_{ij}(n) > E_{ij}(n-1) \\ 0 & other \end{cases}$ $E_{ij}(n) = e^{-\alpha_E} E_{ij}(n-1) + V_E Y_{ij}(n)$

Figure 5b. Equations of PCNN

The indices i, j refer to the ith and jth, neuron , the α terms is decay constant, S_{ij} is the input stimulus, the V's are two respective potentials, M and W are the two synaptic weight sets, and the Y terms refer to the output of neurons from the previous iteration n-1. β is the linking strength of the two components. The state U is compared to a dynamic threshold E to form the output Y of pixel (i, j). As summarized in fig. 5b, when a neuron fires (Y > E), the threshold increases by a large constant amount VF. The neuron is thus prevented from firing for a while, until E decays (according to the decay constant e- α) sufficiently for the value of Y to exceed E once again. To calculate their current values, the threshold, and both the feeding and linking compartments retain a memory of their previous state.

VL -- Magnification factor in the coupling connection area;

aL -- Decay time constant in the coupling connection area;

VF -- Magnification factor in the feedback input area;

αF -- Decay time constant in the feedback input area;

VE-- Magnification factor of the dynamic threshold gate E;

αE-- Decay time constant of the dynamic threshold gate E;

M -- Connection matrix in the coupling connection area;

W -- Connection matrix in the feedback input area;

 β -- The internal activities of the connecting factor;

Sij-- is the external stimulate of current neural cell, commonly is gray value of input image each pixel;

G. Advantages of PCNN

Neural Networks (NN) have been widely employed in face recognition applications. It is feasible for classification and results in similar or higher accuracies from fewer training samples. It has some advantages over traditional classifiers due to two important characteristics: their non-parametric nature and the non-Gaussian distribution assumption. PCNN is called the 3rd generation NN, since it has the following advantages: (i) global optimal approximation characteristic and favorable classification capability; (ii) rapid convergence of learning procedure; (iii) an optimal network to accomplish the mapping unction in the feed-forward; (iv) no need to pre-train. In this article, we present a novel feature extraction approach based on PCNN [14]. PCNN has been used for various applications, such as image feature selection [15, 16] and image segmentation [17] etc.

H. Feature extraction using PCNN

An input gray-scale image is composed of $m \times n$ pixels. This image can be represented as an array of $m \times n$ normalized intensity values. Then the array is fed in at the $m \times n$ inputs of the PCNN. PCNN could put out a series of binary images which correspond to 2D distribution of different gray levels with suitable parameters when a grayscale image inputs it. The 1D temporal series can be derived from the binary images to form the invariable and unique feature representation for classification. PCNN calculated the entropy of each binary image in each iterative procedure at n time and form a temporal

series. The images with different texture would cause the distinct disparities in the number of firing neurons and correspondent firing time sequence because of the difference of gray level distribution. So the output images are varying with each input grayscale image and the correspondent entropy time-series.

Decay time constant: $\alpha_L = \alpha_E = 1$ $\alpha_F = 0.1$ Magnification factor: $V_L = 0.2 V_F = 0.5 V_E = 20$ Connection matrix: $\mathbf{M} = \mathbf{W} = \begin{vmatrix} 0.5 & 1 & 0.5 \\ 1 & 0 & 1 \\ 0.5 & 1 & 0.5 \end{vmatrix}$

Feature vector: $V_i = -P_1 \log_2 P_1 - P_0 \log_2 P_0$; value of i is the iterative times of PCNN; P_1 is probability of value **1** in the put matrix Y; P_0 is probability of value **0** in the put matrix Y.



Figure 6. PCNN features for Mammogram mdb023

VI. GENETIC ALGORITHM FOR FEATURE SELECTION

In this chapter, textural matrices such as GLCM, GLDM, SRDM, GLRLM and PCNN are created for each mammogram image. For each defined distance and direction the Haralick features are extracted for the mammogram images. A single feature value for all the images is considered the initial population string for Genetic Algorithm. An optimum value is computed for each individual feature set. The feature set, which selects the maximum value among other features, is selected for classification. Finally, the algorithm selects the four optimum features from the set of fourteen features. Only the selected features are used for classification.

From the population of the individual feature set, the fitness value is calculated for each feature using the fitness function (1/1+Pi), where Pi is the feature value. Then the probability of each feature value is calculated. And the

cumulative probability is compared for each feature value. Then a random number between zero and one is generated for each feature value. If the cumulative probability value for a feature is higher than the random number, then the feature selection count is incremented by one. This procedure is repeated for the number of times equal to the population size. Next, the population is reproduced with the feature values whose selection count is greater than zero. Each feature is copied into the reproduced population corresponding to the number of times it has been selected. For example, if a selection count for a feature is two, then that feature will be copied two times in the reproduced population.

After reproduction the single point crossover operation is performed on population strings depending upon the crossover probability (Pc). The Pc ranges from zero and one. In the single point crossover operation, initially the pair of population strings is randomly selected for matting. And a random bit position is selected for each pair.

The bits available after the random bit position are exchanged between the population strings in the pair. Thus the matting is performed to create another population set with different values. Next, the mutation operator is applied to the matted population strings depending upon the mutation probability (Pm), where Pm is a small number ranging from zero and one. In mutation, a random bit position is selected from the population. If the bit value is one in that position it is flipped to zero; else it is changed to one. The population now contains a new set of strings for the next population. The next iteration is performed with the new population of strings. This procedure is repeated 30-200 times. Finally the maximum value from the recent population is returned as optimum value of the feature set. Fig. 7 shows the algorithm of feature selection using GA.

| $P_i \leftarrow$ feature values. |
|--|
| $F_i = 1/(1+P_i)$, { Fitness values} |
| Calculate the probability and cumulative probability. CP |
| Reproduction |
| $r \leftarrow random()$ |
| if $(CP_i > r)$ count=count + 1 for CP _i |
| Repeat the steps (a) and (b) for all the population strings. |
| If count=0, then delete that P _i |
| Reproduce the population by copying the selected strings with the corresponding number of times it has |
| been selected. |
| Crossover |
| $r \leftarrow random()$ |
| S select the pair of strings for matting randomly |
| if $(r > P_c)$ $k \leftarrow$ random bit position |
| interchange the bits after k th position in parent1 and parent2 |
| repeat this step for all the pairs. |
| Mutation |
| $r \leftarrow random()$ |
| if $(r \ge P_m)$, k \leftarrow random bit position |
| complement the value of the k th bit |
| repeat this steps for all the strings. |
| $P_{new} \leftarrow population after reproduction, crossover and mutation.$ |
| $P_i \leftarrow P_{new}$ |
| Goto Step 2 |
| |

Figure 7. Algorithm of feature selection using GA

The four selected Haralick features: correlation, sum average, difference variance, maximal correlation coefficient are used for classification. The four features are selected in runlength features; these are short run emphasis, long run emphasis, gray level non-uniformity, run length non-uniformity. And the four features whichever the maximum value from the optimum value of the PCNN feature set are selected by genetic algorithm.

VII. FUZZY C-MEANS (FCM) CLUSTERING

The Fuzzy C-means clustering is used to classify the dataset. The dataset is clustered into different number of clusters such as 2, 3, 4, 5, and 6. The fuzzy index value of clusters, the mean distance of clusters and the squared errors of clusters are evaluates the classification performance. The classification performance is high, if the fuzzy index value of clusters and the mean distance of clusters becomes high, and the squared errors of clusters becomes low[18, 19, 20].

Fuzzy C-means clustering is a process designed to assign each sample to a cluster based on cluster membership. The algorithm is based on iterative minimization of the following function:

$$J(U,V) = \sum_{i=1}^{c} \sum_{k=1}^{n} u_{ik}^{m} |x_{k} - v_{i}|^{2}$$
(3)

where, $x_{1,...,} x_n$ are *n* data sample vectors; $V = \{v_1,...,v_c\}$ are cluster centers;

 $U = [u_{ik}]$ is a c x n matrix, where u_{ik} is the i th membership value of the k th input sample x_k , and the membership values satisfy the following conditions

$$0 \le u_{ik} \le 1 \qquad i=1,2,...,n$$

$$\sum_{i=1}^{c} u_{ik} = 1 \qquad k=1,2,...,n$$

$$<\sum_{k=1}^{n} u_{ik} < n \qquad i=1,2,...,c$$

 $m \in [1, \infty]$ is an exponent weight factor.

VIII. EXPERIMENTAL RESULTS AND DISCUSSION

The Experiment procedure is given as follows:

A. Acquire mammogram image from MIAS data base

In this experiment, we use mammogram images, which contains normal, benign, and malignant mammogram images.

B. Image Enhancement

The mammogram images should be pre-processed. In this research, median filter has been applied for de-noising the images [25]. The 8-neighborhood connected component labelling method is used for removing artifacts and pectoral muscles [24].

C. Image segmentation

The suspicious region or micro calcifications are segmented using Darwinian Particle Swarm Optimization for mammogram images [21].

D. Extracted features

Generate the texture description matrices GLCM, GLDM, SRDM from the segmented image and extracted 14 Haralick features. Generate GLRLM and extracted 7 runlength features. The PCNN features are extracted from entropy time series formula. The features are normalized between 0 and 1 using the equation (1).

$$y(x) = \frac{y - y_{\min}}{y_{\max} - y_{\min}}$$
(5)

where y(x) is the normalized value (between 0 and 1), y is the original value, y_{min} is the minimum allowed value, y_{max} is the maximum allowed value

E. Feature selection

The features are selected from genetic algorithm, the four selected haralick features : correlation , sum average, difference variance, maximal correlation coefficient are used for classification. The four features are selected in runlength features, these are short run emphasis, long run emphasis, gray level non-uniformity, run length non-uniformity [22]. And the four features whichever the maximum value from the optimum value of the PCNN feature set are selected by genetic algorithm.

F. Classification and analysis

Fuzzy C-means clustering algorithm is used to classify the features and analyses according to the fuzzy index value, mean distance among cluster centers, and classification squared error value of clusters [23].

G. Performance Analysis

The Euclidian distance based linkage dendrogram shows the classification performance of various feature extraction method. Fig. 9 shows the Classification performance of GLDM method. Fig. 10 shows the Classification performance of SRDM method, Fig. 11 shows the Classification performance of GLRLM method and Fig. 12 shows the Classification performance of PCNN method The Euclidean distance formula is as follows.

Distance
$$(x_i, x_j) = \sqrt{\sum_{k=1}^{n} (x_{ik} - x_{jk})^2}$$
 (6)

for all i, j = 1, 2, ..., m

(4)



TABLE IV. COPHENETIC CORRELATION VALUE OF GLCM, GLDM, SRDM, GLRLM, AND PCNN METHODS

| GLCM | GLDM | SRDM | GLRLM | PCNN |
|--------|--------|--------|--------|--------|
| 0.9369 | 0.7836 | 0.4950 | 0.7493 | 0.8011 |

Fig. 13 shows the comparisons for classification performance of GLCM, GLDM, SRDM, GLRLM, and PCNN methods. According to the cophenetic correlation coefficient. The magnitude of this value should be very close to 1 for a high-quality solution. In our experiment GLCM and PCNN methods are close to the value 1, hence the GLCM and PCNN performs more efficient than other methods.

H. Performance Cophenetic correlation

The cophenet function measures the distortion of this classification, indicating how readily the data fits into the structure suggested by the classification. The output value, c, is the cophenetic correlation coefficient. The magnitude of

this value should be very close to 1 for a high-quality solution. This measure can be used to compare alternative cluster solutions obtained using different algorithms. The cophenetic correlation between Z(:,3) and Y is defined as

$$C = \frac{\sum_{i < j} (Y_{ij} - y)(Z_{ij} - z)}{\sqrt{\sum_{i < j} (Y_{ij} - y)^2 \sum_{i < j} (Z_{ij} - z)^2}}$$
(7)

where: Y_{ij} is the distance between objects i and j in Y. Z_{ij} is the distance between objects i and j in Z. y and z are the average of Y and Z, respectively.

I. Fuzzy classification index value

To Determine the classification performance of clusters, the fuzzy classification index value, Fc was calculated based on different numbers of partition groups in the distract partition λ -cutsets. The classification performance be high, if the classification index value of clusters become high [23]. These Fc values were calculated using the following formula [13].

Fc =
$$\frac{(m-c) \Sigma^{c}{}_{n=1}m_{n}(mx_{n}-mx) (mx_{n}-mx)^{T}}{(c-1) \Sigma^{c}{}_{n=1} \Sigma^{c}{}_{r=1} (x_{r}-mx_{n}) (x_{r}-mx_{n})^{T}}$$

where m is the total number of objects; c is the number of partition groups; mn is total number of samples in the nth partition group; r is the new sequence number in the nth partition group; mxn is the mean value of objects in the nth partition group; mx is the mean value of all the objects; (mxn-mx)T and (xr-mxn)T be the transpose of (mxn-mx) and (xr-mxn) respectively.

Fig. 14 shows the classification performance of Cluster vs index value method and Table V shows the Classification index value



Figure 14. The Classification performance of cluster vs index value of GLCM, GLDM, SRDM, GLRLM and PCNN methods

J. Centroid

The Centroid is computed by taking mean value of each Cluster.

$$C_i = (\Sigma_{r=1}^n X_r)/n;$$

where r is the new sequence number in the i th partition group.

The dataset is classified into different number of clusters such as 2, 3, 4, 5 and 6. The mean distance of clusters centers is evaluated for each data set. Fig. 15 shows the classification performance against mean distance of clusters centres and Table VI shows the Mean distance Value among clusters centers of GLCM, GLDM, SRDM, GLRLM. and PCNN methods.

TABLE V. THE CLASSIFICATION PERFORMANCE INDEX VALUE OF GLCM, GLDM, SRDM, GLRLM AND PCNN METHODS

| Cluster | GLCM | GLDM | SRDM | GLRLM | PCNN |
|---------|--------|--------|--------|--------|--------|
| 2 | 14.219 | 17.071 | 38.125 | 52.420 | 112.42 |
| 3 | 36.482 | 16.424 | 37.881 | 56.006 | 144.44 |
| 4 | 30.985 | 14.596 | 37.378 | 51.706 | 194.19 |
| 5 | 27.386 | 14.534 | 37.237 | 49.319 | 197.12 |
| 6 | 23.748 | 11.598 | 35.642 | 42.414 | 179.52 |

(8)



Figure 15: The Classification performance of cetroid vs Mean Distance of GLCM, GLDM, SRDM, GLRLM. and PCNN methods

TABLE VI. THE CLASSIFICATION PERFORMANCE MEAN DISTANCE VALUE OF CENTROID OF GLCM, GLDM, SRDM, GLRLM AND PCNN METHODS.

| Cluster | GLCM | GLDM | SRDM | GLRLM | PCNN |
|---------|--------|--------|--------|--------|--------|
| 2 | 0.2288 | 0.3622 | 0.5944 | 0.7576 | 0.7430 |
| 3 | 0.6869 | 0.3639 | 0.6015 | 0.7923 | 0.7043 |
| 4 | 0.5813 | 0.3821 | 0.6231 | 0.7132 | 0.9736 |
| 5 | 0.5221 | 0.3894 | 0.6492 | 0.6498 | 0.8539 |
| 6 | 0.4845 | 0.4048 | 0.6506 | 0.7669 | 0.7888 |

K. Squared Errors

The squared errors among clusters are also calculated for the classification performance. The classification performance be high, if the squared error among clusters becomes low [24]. Fig. 16 shows the classification performance of GLCM, GLDM, SRDM, GLRLM and PCNN methods against squared error of clusters. Table VII shows the squared error Value of GLCM, GLDM, SRDM, GLRLM and PCNN methods.



TABLE VII. THE CLASSIFICATION PERFORMANCE SQUARED ERROR VALUE OF GLCM, GLDM, SRDM, GLRLM AND PCNN METHODS

| Cluster | GLCM | GLDM | SRDM | GLRLM | PCNN |
|---------|--------|--------|--------|--------|--------|
| 2 | 5.4846 | 6.4045 | 6.0276 | 5.2762 | 2.7010 |
| 3 | 2.7853 | 5.1105 | 4.1406 | 3.2627 | 1.2631 |
| 4 | 2.3535 | 4.4480 | 3.1460 | 2.5247 | 0.6606 |
| 5 | 2.0701 | 3.7882 | 2.5093 | 2.0502 | 0.4874 |
| 6 | 1.9222 | 3.7455 | 2.1415 | 1.8967 | 0.4218 |

Figure 16: The Classification performance of cetroid vs Squared Error of GLCM, GLDM, SRDM, GLRLM. and PCNN methods

L. Results

In this experimental work, we consider four types of texture description methods such as GLCM, GLDM, SRDM, GLRLM and a PCNN feature extraction method are analysed in the digital mammogram image. The features are classified into different number of clusters such as 2, 3, 4, 5 and 6 using FCM method. In the evaluation, the classification performance of the PCNN feature extraction method performs more efficient than other methods since the fuzzy classification index is very high, the mean distance of clusters value is high and squared error is very low than other methods.

IX. CONCLUSION

Textural features are extracted for classification of micro-calcifications. In this work we consider four types of texture description methods such as GLCM, GLDM, SRDM, GLRLM and PCNN feature extraction method are analysed in the mammogram image. The features are classified into different number of clusters such as 2, 3, 4, 5 and 6 using FCM method. The fuzzy classification index, the mean distance of clusters and squared error evaluate are used to classification performance. It was observed that the performance of classification based on PCNN. PCNN feature extraction method performs more efficient than other methods.

ACKNOWLEDGMENT

The second author immensely acknowledges the UGC, New Delhi for partial financial assistance under UGC-SAP (DRS) Grant No. F.3-50/2011.

The first and fourth authors immensely acknowledge the partial financial assistance under University Research Fellowship, Periyar University, Salem.

REFERENCES

- [1] Breast Cancer in Men. A complete patient's guide. http://www.breastdoctor.com/breast/men/cancer.htm.
- [2] Breast Cancer in Men. Male breast cancer information center. http://interact.withus.com/interact/mbc/.
- [3] C. Velayutham, and K. Thangavel, "Unsupervised Feature Selection in Digital Mammogram Image Using Rough Set Based Entropy Measure", Proceedings of the IEEE World Congress on Information and Communication Technologies, pp. 632-637, 2011.
- [4] C. Velayutham, and K. Thangavel, "Entropy based unsupervised Feature Selection indigital mammogram image using rough set theory", International Journal of Computational Biology and Drug Design, vol. 5, no. 1, pp.16–34, 2012.
- [5] http://www.wiau.man.ac.uk/services/MIAS/MIASweb.html.
- [6] Jiawei Han and Micheline Kamber. Data Mining Concepts and Techniques. Morgan Kaufmann, 2001.
- [7] Rafael C. Gonzalez and Richard E. Woods. Digital Image Processing, 2nd edition. Addison-Wesley, 1993.
- [8] Thangavel, K., and Karnan, M.; "CAD system for Preprocessing and Enhancement of Digital Mammograms," International Journal on Graphics Vision and Image Processing, vol. 5, no. 9, pp. 69-74, 2005.
- [9] Nadler, M., and Smith, E.P.: "Pattern Recognition Engineering," New York: Wiley, 1993.
- [10] Haralick et al., (1973) Texture Features for Image Classification. IEEE Trans Syst. Man Cyb., Vol. SMC-3, no.6, pp. 610-621.
- [11] R.M. Haralick Haralick (1979) Statistical and Structural Approaches to Texture, Proc 4th Int. Joint Conf. Pattern Recog., pp. 45-60.
- [12] Gulsrud, T.O.: "Texture analysis of Digital Mammograms," Ph.D. Thesis, Aalborg University, Stavanger, USA, pp: 30-32, 2000.
- [13] H. A. Abbass "An agent based approach to 3-SAT using marriage in honey-bees optimization", International Journal of Knowledge-Based Intelligent Engineering Systems (KES), vol. 6, no. 2, pp. 1-8.
- [14] YuDong Zhang and LeNan Wu, "Pattern Recognition via PCNN and Tsallis Entropy", Sensors 2008, 8, 7518-7529; DOI: 10.3390/s8117518. ISSN 1424-8220, www.mdpi.com/journal/sensors, 2008.
- [15] J.L. Johnson, "Pulse-coupled Neural Nets: Translation, Rotation, Scale, Distortion and Intensity Signal Invariance for Image", Patt. Recogn. Lett., 33, 6239-6253, 1994.
- [16] Trong-Thuc Hoang, Ngoc-Hung Nguyen, Xuan -Thuan Nguyen and Trong-Tu Bui, "A Real-time Image Feature Extraction Using Pulse-Coupled Neural Network", International Journal of Emerging Treads and Technology in Computer Science, Volume 1, Issue 3, September – October 2012.
- [17] G.Kuntimad,; H.S. Ranganath, "Perfect Image Segmentation Using Pulse Coupled Neural Networks", IEEE Trans. Neural Netw. 10, 591-598, 1999.
- [18] Periklis Andritsor(2002), "data clustering Techniques" pp. 1-34.
- [19] Bezdek.J(1974) "Numerical taxonomy with fuzzy sets" J.math.Biol vol. 1 pp. 57 71.
- [20] Hala S. Own, "Aboul Ella Hassanien Rough Wavelet Hybrid Image Classification Scheme, Journal of Convergence Information Technology vol. 3, no. 4, pp. 65-75, December 2008.
- [21] Pedram Ghamisi, Micael S. Couceiro, Jón Atli Benediktsson and Nuno M.F. Ferreira "An efficient method for segmentation of images based on fractional calculus and natural selection", Expert Systems with Applications 39 (2012).
- [22] J.M. Steppe "Feature and Model Selection in Feed forward Neural Networks," Ph.D Dissertation, Air Force Institute of Technology, Air University USA, 1994.
- [23] Bezdek, J.C., and Pal, S.K.: "Fuzzy Models for Pattern Recognition," IEEE Press, New York, 1992.
- [24] R. Subash Chandra Boss, K. Thangavel and D. Arul Pon Daniel, "Automatic Mammogram image Breast Region Extraction and Removal of Pectoral Muscle", International Journal of Scientific & Engineering Research, vol. 4, issue2, February-2013.
- [25] R. Subash Chandra Boss, K. Thangavel, and D. Arul Pon Daniel. "Mammogram image segmentation using fuzzy clustering" Pattern Recognition, Informatics and Medical Engineering (PRIME), 2012 International Conference on. IEEE, 2012.