

Microcalcifications Detection Using PFCM and ANN

A. Vega-Corona², J. Quintanilla-Domínguez^{1,3,*}, B. Ojeda-Magaña^{1,2},
M.G. Cortina-Januchs^{1,3}, A. Marcano-Cedeño¹, R. Ruelas², and D. Andina¹

¹ Group for Automation in Signals and Communications GASC
Technical University of Madrid, 28040 Madrid, Spain

{joelq, januchs}@salamanca.ugto.mx, {a.marcano, d.andina}@gc.ssr.upm.es

² Computational Intelligence Laboratory LABINCO-DICIS,
University of Guanajuato. 36885 Salamanca Guanajuato, Mexico
tono@salamanca.ugto.mx

³ Department of Projects Engineering DIP-CUCEI,
University of Guadalajara. 45101 Zapopan Jalisco, Mexico
benojed@hotmail.com, rruelas@newton.dip.udg.mx

Abstract. This work presents a method to detect *Microcalcifications* in *Regions of Interest* from digitized mammograms. The method is based mainly on the combination of *Image Processing*, *Pattern Recognition* and *Artificial Intelligence*. The *Top-Hat transform* is a technique based on mathematical morphology operations that, in this work is used to perform contrast enhancement of microcalcifications in the region of interest. In order to find more or less homogeneous regions in the image, we apply a novel *image sub-segmentation* technique based on *Possibilistic Fuzzy c-Means* clustering algorithm. From the original region of interest we extract two window-based features, *Mean* and *Deviation Standard*, which will be used in a classifier based on a *Artificial Neural Network* in order to identify microcalcifications. Our results show that the proposed method is a good alternative in the stage of microcalcifications detection, because this stage is an important part of the early *Breast Cancer* detection.

Keywords: Microcalcifications detection and classification, Top-Hat transform, Possibilistic Fuzzy c-Means, Artificial Neural Networks.

1 Introduction

Breast cancer is the most common types of cancer among women all over the world. Early detection continues being a key piece to improve the prognosis and

* The authors wish to thank the National Council for Science and Technology (CONACYT), the Group of Automation in Signal and Communications (GASC) of the Technical University of Madrid, Computational Intelligence Laboratory LABINCO-DICIS of University of Guanajuato and Department of Projects Engineering DIP-CUCEI of University of Guadalajara.

survival of breast cancer cases [1]. Currently the most reliable and practical method for early detection and screening of breast cancer is mammography. The presence of *Microcalcifications (MCs)* clusters has been considered as a very important indicator of malignant types of breast cancer and its detection and classification are important to prevent and treat the disease. However, the detection and classification of MCs continue being a hard work due to that, in mammograms there is a poor contrast between MCs and the tissue around them. The MCs are tiny deposits of calcium in breast tissue. The MCs appear in the small clusters of a few pixels with relatively high intensity and closed contours compared with their neighboring pixels. Individual MCs are sometimes difficult to detect because of the surrounding breast tissue, their variation in shape, orientation, brightness and diameter size [11].

This work presents a method based on *Image Processing, Pattern Recognition* and *Artificial Intelligence* for identification of MCs in digitized mammogram images. The method consist in: Image selection, image enhancement by morphological *Top-Hat transform*, data clustering and labeling by a partitional clustering algorithm such as *Possibilistic Fuzzy c-Means (PFCM)*, feature extraction based on window-based features such as *mean* and *standard deviation* and finally a classifier based on an *Artificial Neural Network (ANN)*.

2 ROI Selection

The proposed method has been trained and tested using a set of mammograms extracted from a mini-Mammographic database provided by Mammographic Image Analysis Society (MIAS) [10]. Each mammogram from the database is a 1024×1024 pixels and with a spatial resolution of $200 \mu\text{m}/\text{pixel}$. These mammograms have been reviewed by an expert radiologist and all the abnormalities have been identified and classified. To the place where these abnormalities such as, MCs, have been located is known as, *Region of Interest (ROI)*. In this work, the ROI images with a size of 256×256 were used.

3 Detection of Microcalcifications

3.1 Enhancement of Microcalcifications

In the past several years methodologies to detect and / or classify MCs have been developed, but the detection of MCs continues being a difficult task due to some factors, such as, size, shape, low contrast and low visibility from their surrounding as well as the distribution with respect to their fuzzy morphology [4]. The contrast can be defined as the difference in intensity between an image structure and its background. By combining of morphological operations, several image processing tasks can be performed, but in this work we used morphological operations to achieve contrast enhancement.

With the aim to improve the contrast between the MCs and background in the ROI image, we use a *morphological contrast enhancement*. This morphological contrast enhancement, is an image enhancement technique based on Mathematical Morphology operations known as Top-Hat and Bottom-Hat transforms.

But, in this work we will consider only the Top-Hat transform. In previous works such as, [9], [12], this technique was used to obtain satisfactory results in the MCs detection stage.

A Top-Hat is a residual filter which preserves those features in an image that can fit inside the structuring element and removes those that cannot, in other words the Top-Hat transform is used to segment objects that differ in brightness from the surrounding background in images with uneven background intensity. The Top-Hat transform is defined by the following equation:

$$I_T = I - [(I \ominus SE) \oplus SE] \quad (1)$$

where, I is the input ROI image, I_T is the transformed ROI image, SE is the structuring element, \ominus represents morphological erosion operation, \oplus represents morphological dilation operation and $-$ image subtraction operation. $[(I \ominus SE) \oplus SE]$ is also known as the morphological opening operation.

3.2 Segmentation of Microcalcifications by PFCM

Digital image segmentation has been considered the most important intermediate step in image processing to extract a semantic meaning of the pixels. The main objective of clustering process in order to image segmentation is to find groups of pixels with similar features such as, gray level intensity, or more or less homogeneous groups. The similarity is evaluated according to a distance measure between the pixels and the prototypes of groups, and each pixel is assigned to the group with the nearest or most similar prototype [7].

In this work we used a novel method to segment MCs in ROI images, this method was called image sub-segmentation and it is based on *Possibilistic Fuzzy c-Means (PFCM)* clustering algorithm. This method was proposed in our previous work [7].

The PFCM is a more recent partitional clustering algorithm, which has the advantages of the Fuzzy c -Means (FCM) as well as the Possibilistic c -Means (PCM). The FCM has a constraint that makes it very sensitive to outliers. To solve the problem of constraint of the FCM, Krisnapuram and Keller [6] have developed the clustering algorithm PCM, which allows us to identify the degree of typicality which has a data point with respect to the group that belongs. The PCM has a problem, sometimes the prototypes of clusters coincide generating erroneous partitions of the feature space, for this reason it was not so successful. Pal *et al.*[8] to solve the problems of the clustering algorithms as FCM (outlier sensitivity) and PCM (coincident clusters) proposed a hybridized PFCM clustering model. This algorithm has four parameters (m, η, a y b), where the values of a and b represent a relative importance of membership and typicality values in the computation of prototypes respectively. The parameters m and η

represent an absolute weight of membership value and typicality value, respectively. In order to reduce the effect of outliers, it is advised to set $b > a$ and $m > \eta$.

Proposed approach for detection of MCs by sub-segmentation

1. Get the data vector.
2. Assign a value to the parameters (a, b, m, η) .
3. Segment the image taking into account the number of more representative regions, in this case two: Suspicious region with presence of MCs (S_1) and Normal Tissue (S_2), this region we will consider without the presence of MCs.
4. Run the PFCM algorithm to get:
 - The membership matrix U .
 - The typicality matrix T .
5. Get the maximum typicality value for each pixel.

$$T_{max} = \max_i [t_{ik}], \quad i = 1, \dots, c. \tag{2}$$

6. Select a value for the threshold α .
7. With α and the T_{max} matrix, separate all the pixels into two sub-matrices (T_1, T_2), with the first matrix:

$$T_1 = T_{max} \geq \alpha \tag{3}$$

containing the typical pixels of both regions ($Stypical_1$) and ($Stypical_2$), and the second matrix

$$T_2 = T_{max} < \alpha \tag{4}$$

containing the atypical pixels of both regions ($Satypical_1$) and ($Satypical_2$), in this case the atypical pixels are of the most interest, especially the *atypical* pixels of the (S_1)

8. From the labelled pixels z_k of the T_1 sub-matrix the following sub-regions can be generated

$$T_1 = Stypical_{1, \dots, Stypical_i}, \quad i = 1, \dots, c. \tag{5}$$

and from the T_2 sub-matrix

$$T_2 = Satypical_{1+i, \dots, Satypical_{2i}}, \quad i = 1, \dots, c. \tag{6}$$

such that each region $S_i, i = 1, \dots, c$ is defined by

$$S_i = Stypical_i \cup Satypical_{i+c}. \tag{7}$$

9. Select the sub-matrix T_1 or T_2 of interest for the corresponding analysis. In the case of this work, T_2 is the sub-matrix of interest.
-

4 Classification of Microcalcifications

4.1 Feature Extraction

MCs appear on mammograms as bright spots. These bright spots are small regions with intensity values higher than their surroundings or background. Spatial domain features includes both shape-related features and window-based features. In this work we applied window-based features [4] [5]. These features are *mean* and *standard deviation*, extracted from the original ROI images within a rectangular window size $n \times m$ centered into position (i, j) . The purpose of applying these characteristics is to distinguish the pixels that correspond to possible MCs of the pixels that correspond to background. The mean and standard deviation are defined by:

$$I_{\mu}(i, j) = \frac{1}{n \times m} \sum_{i=1}^n \sum_{j=1}^m f(i, j) \quad (8)$$

$$I_{\sigma}(i, j) = \left[\frac{1}{n \times m} \sum_{i=1}^n \sum_{j=1}^m (f(i, j) - I_{\mu}(i, j))^2 \right]^{1/2} \quad (9)$$

where, I_{μ} , I_{σ} and $f(i, j)$ represent the mean, deviation standard and the gray level value of a pixel located in (i, j) respectively.

4.2 Classification of Microcalcifications by ANN

Artificial Neural Networks (ANNs) are biologically inspired networks based on the neuron organization and decision making process in human brain [2]. In other words, it is the mathematical model of the brain. ANNs are used in a wide variety of data processing applications where real-time data analysis and information extraction are required. One advantage of ANNs approach is that most of the intense computation takes place during the training process. Once ANNs are trained for a particular task, operation is relatively fast and unknown samples can be rapidly identified in the field. Classification is one of the most frequently encountered decision making task of human activity. A classification problem occurs when an object needs to be assigned into a predefined group or class based on a number of observed attributes related to that object. In this work, we proposed a classifier based on ANN in order to classify the patterns such as MCs class or Normal Tissue class. For this purpose, we applied Multilayer Perceptron (MLP). The MLP is the most used in many practical applications, such as in pattern recognition applications. The functionality of the MLP topology is determined by a learning algorithm, the Back Propagation (BP) [3], based on the method of steepest descent. In the process of upgrading the connection weights, it's the most commonly used algorithm by the ANN scientific community.

5 Methodology and Results

In this work we consider applying our method to each ROI individually in order to show the obtained results by means of a segmented image. We selected ten ROI images from mammograms with dense tissue and the presence of MCs.

Morphological Enhancement. With the aim to detect objects that differ in brightness from the surrounding background, the Top-Hat transform is used. In this case the Top-Hat transform is used to increase the contrast between the MCs and background. Then, we apply the same SE at different sizes, 3×3 , 5×5 , 7×7 , to perform Top-Hat transform. The SE used in this work is a flat *disk-shaped* SE . Fig. 1 shows the original ROI images processed by Top-Hat transform.

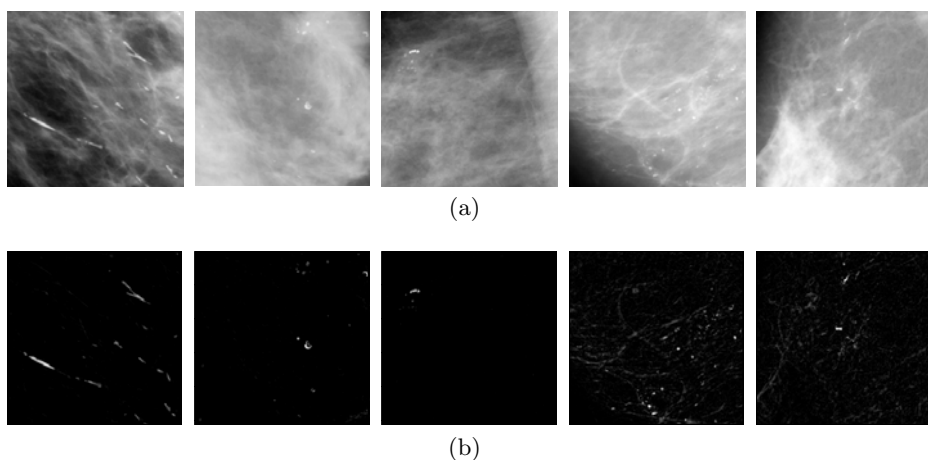


Fig. 1. (a)Original ROIs images (mdb148, mdb170, mdb219, mdb245, mdb249). (b) ROIs images processed by *Top-Hat* transform.

Segmentation of Microcalcifications by PFCM. We consider the ROI images processed by Top-Hat transform to build a *data vector*. In this data vector we apply clustering by PFCM to obtain a label for each pattern belonging to each group of the partition of feature space.

From the point of view of image, with the sub-segmentation we obtain a better identification of possible pixels belonging to MCs, because we found the most atypical pixels of each regions, this is, all those pixels that are away from their prototype, these pixels regularly are very small group with intensity values higher than their surroundings and are very different from the rest of pixels in the ROI image.

Fig. 2 shows the process of image sub-segmentation applying the PFCM, Fig. 2(b) shows suspicious region with presence of MCs and Normal Tissue region, Fig. 2(c) shows each region sub-divided into 4 regions $Stypical_1$, $Satypical_3$ and $Stypical_2$, $Satypical_4$, Fig. 2(d) shows the $atypical_3$ data (pixels) of the

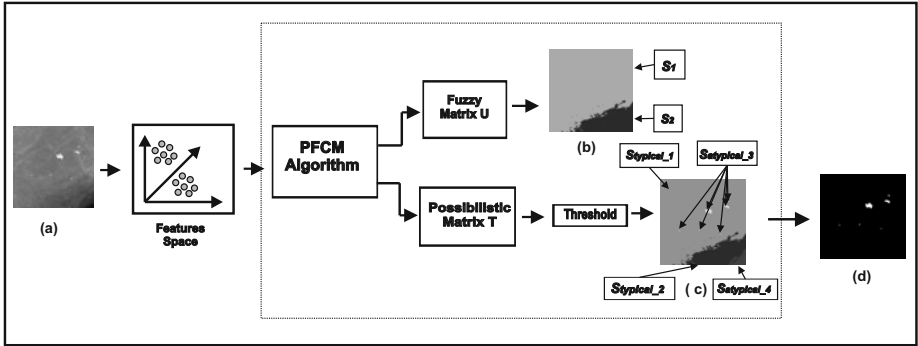


Fig. 2. Process of image sub-segmentation by PFCM

(S_1), which is considered as MCs region, after applying a threshold (α) value. The process of sub-segmentation was performed to 10 ROIs image with the following initial parameters, $a = 1, b = 2, m = 2, \eta = 2, \alpha = 0.6$, and we obtained the following values: The number of pixels assigned to the class MCs were 3105 and the number of pixels assigned to the class Normal Tissue were 652255.

Feature Extraction. Two window-based features such as, mean and standard deviation were extracted. These features are extracted from original ROI images within a rectangular windows, in this work we used three different pixel block windows of size $3 \times 3, 5 \times 5, 7 \times 7$. In our work, each image obtained by window-based features, are considered as a features to generate a set of patterns that represent MCs and Normal Tissue. This set of patterns we called feature vector (FV). Each of the images used in this work, we know *a priori* that, there are pixels belonging to MCs or Normal Tissue. This FV is considered as a input vector for the classifier. The FV were formed as follow:

$$FV = [I_{\mu_{3 \times 3}}, I_{\sigma_{3 \times 3}}, I_{\mu_{5 \times 5}}, I_{\sigma_{5 \times 5}}, I_{\mu_{7 \times 7}}, I_{\sigma_{7 \times 7}}] \tag{10}$$

The labels of the two classes of the FV were obtained in the previous step by means of sub-segmentation by PFCM. Due to the large amount of patterns of the class that does not belong to MCs respect to the number of patterns that belongs to the class of MCs a balancing was performed, then we obtain two subsets with 3105 patterns to the class MCs and 31050 patterns to the class Normal Tissue.

Classification of Microcalcifications by ANN. In this stage of our method, we used a classifier based on ANN in order to classify the patterns such as Normal Tissue class or MCs class. For this purpose, we applied MLP classifier. We apply different network structures, of which the best obtained results were of the following structure and parameters:

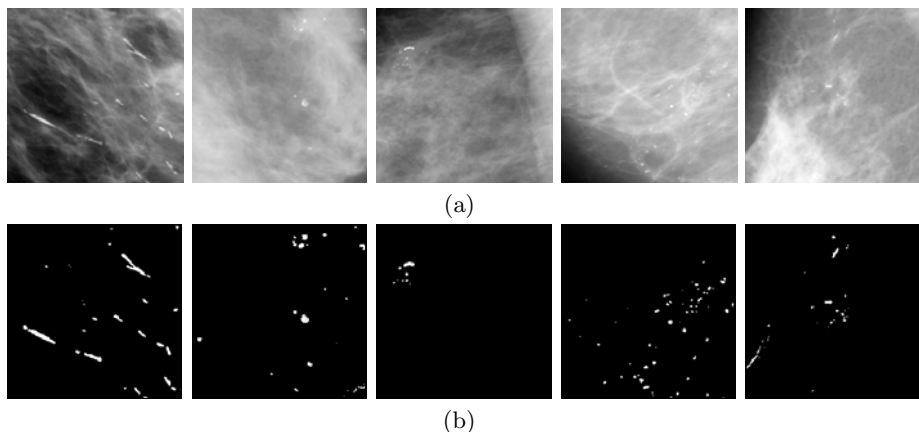


Fig. 3. (a)Original ROIs images. (b)MCs detection using our method.

Table 1. Number of patterns using for training and testing by each classifier

	Numbers of Sample		Total
	Training	Testing	
MCs	2489	616	3105
Normal Tissue	24835	6215	31050

Table 2. Confusion matrix and performance of the classifier

Classifier	Desired Results	Output Results		Sensitivity (%)	Specificity (%)	Total Classification Accuracy (%)
		MCs	Normal Tissue			
MLP <i>Structure</i> ₁ 6 : 24 : 1	MCs	605	11	98.21	99.68	99.54
	Normal Tissue	20	6195			

1. Number of input neurons equal to the number of features, for this case 6.
2. Number of hidden layers: 1.
3. Hidden neurons: 24.
4. Output neurons: 1.
5. Learning rate η : 1.
6. Activation function: sigmoidal with value between [0,1].
7. Training conditions, epochs and Mean Squared Error: 250 and 0.001 respectively

We used patterns extracted of the set *FV* in order to train and test our classifier, for this case we used the 80% and 20% of the data randomly selected, respectively. Table 1 shows the distribution of selected data.

Next, we built a confusion matrix to determine the probability of the detection MCs vs. probability of false MCs. Table 2 shows the performance of the classifier presented in this work.

Finally, Fig.3 shows the obtained results of the MCs detection in the ROIs using the method proposed in this work.

6 Conclusions

The experiment results show that the proposed method can locate MCs in an efficient way, moreover the method promises interesting advances in Medical Field. The analysis of the results served as comparison point with other methods used by other authors mentioned in this work. On the other hand, the use of PFCM was very helpful at the stage of MCs segmentation, because one of the main advantages offered by this algorithm, which is to find typical or atypical data of a region based on the established threshold, thus for case of MCs, the atypical data could be of greater interest than typical data of the same region. Finally, the obtained results show that the implemented method was able to detect MCs satisfactorily, fulfilling the goal of this work.

References

1. World health organization (WHO), <http://www.who.int/en/>
2. Andina, D., Pham, D.: Computational intelligence for engineering and manufacturing, 1st edn. Springer, Heidelberg (2007)
3. Basheer, I.A., Hajmeer, M.: Artificial neural networks: fundamentals, computing, design, and application. *Journal of Microbiological Methods* 43(1), 3–31 (2000)
4. Cheng, H.D., Cai, X., Chen, X., Hu, L., Lou, X.: Computer-aided detection and classification of microcalcifications in mammograms: a survey. *Pattern Recognition* 36(12), 2967–2991 (2003)
5. Fu, J.C., Lee, S.K., Wong, S.T.C., Yeh, J.Y., Wang, A.H., Wu, H.K.: Image segmentation feature selection and pattern classification for mammographic microcalcifications. *Computerized Medical Imaging and Graphics* 29(6), 419 (2005)
6. Krishnapuram, R., Keller, J.M.: A possibilistic approach to clustering. *IEEE Transactions on Fuzzy Systems* 1(2), 98–110 (1993)
7. Ojeda-Magaña, B., Quintanilla-Domínguez, J., Ruelas, R., Andina, D.: Images subsegmentation with the pfc clustering algorithm. 7th IEEE International Conference on Industrial Informatics, 499–503 (2009)
8. Pal, N.R., Pal, S.K., Keller, J.M., Bezdek, J.C.: A possibilistic fuzzy c-means clustering algorithm. *IEEE Transactions on Fuzzy Systems* 13(4), 517–530 (2005)
9. Stojić, T., Reljin, B.: Enhancement of microcalcifications in digitized mammograms: Multifractal and mathematical morphology approach. *FME Transactions* 38, 1–9 (2010)
10. Suckling, J., Parker, J., Dance, D.: The mammographic image analysis society digital mammogram database. *Exerpta Medica International Congress Series* 1069, 375–378 (1994)
11. Wei, L., Yang, Y., Nishikawa, R.M.: Microcalcification classification assisted by content-based image retrieval for breast cancer diagnosis. *Pattern Recognition* 42(6), 1126 (2009)
12. Wirth, M., Fraschini, M., Lyon, J.: Contrast enhancement of microcalcifications in mammograms using morphological enhancement and non-flat structuring elements. 17th IEEE Symposium on Computer-Based Medical System, 134–139 (2004)