

RESEARCH

Open Access

# Improvement for detection of microcalcifications through clustering algorithms and artificial neural networks

Joel Quintanilla-Domínguez<sup>1,3\*</sup>, Benjamín Ojeda-Magaña<sup>1,2</sup>, Alexis Marcano-Cedeño<sup>1</sup>, María G Cortina-Januchs<sup>1,3</sup>, Antonio Vega-Corona<sup>3</sup> and Diego Andina<sup>1</sup>

## Abstract

A new method for detecting microcalcifications in regions of interest (ROIs) extracted from digitized mammograms is proposed. The *top-hat* transform is a technique based on mathematical morphology operations and, in this paper, is used to perform contrast enhancement of the microcalcifications. To improve microcalcification detection, a novel *image sub-segmentation* approach based on the *possibilistic fuzzy c-means* algorithm is used. From the original ROIs, window-based features, such as the mean and standard deviation, were extracted; these features were used as an input vector in a classifier. The classifier is based on an *artificial neural network* to identify patterns belonging to microcalcifications and healthy tissue. Our results show that the proposed method is a good alternative for automatically detecting microcalcifications, because this stage is an important part of early breast cancer detection.

**Keywords:** detection of microcalcifications, top-hat transform, possibilistic fuzzy c-means clustering algorithm, artificial neural networks

## 1 Introduction

Breast cancer is one of the most serious types of cancer that affects women around the world. It is also one of the leading causes of mortality in middle-aged and elderly women. The International Agency for Research on Cancer (IARC) estimates that more than 1 million cases of breast cancer occur world-wide each year, with some 580,000 cases occurring in developed countries and the remainder in developing countries. The risk of a woman developing breast cancer during her lifetime is approximately 11% [1]. Early detection of breast cancer is of vital importance to successful treatment, with the main goal of increasing the probability of survival for patients. Currently, the most reliable and practical method for early detection and screening of breast cancer is mammography. Microcalcifications (MCs) can be an important early sign of breast cancer; they appear as bright spots of calcium deposits. Individual MCs are sometimes difficult to detect because of the surrounding

breast tissue and variations in shape, orientation, brightness and diameter [2]. MCs are potential primary indicators of malignant types of breast cancer. Therefore, their detection can be important in preventing and treating the disease. However, it is still difficult to detect all MCs in mammograms because of the poor contrast against the tissue that surrounds them.

Many methodologies have been presented by different authors to detect the presence of MCs in mammograms. These methodologies involve image processing techniques, pattern recognition methods and artificial intelligence approaches. Vega-Corona et al. [3] proposed a method for detecting MCs in digitized mammograms. The method consists of image enhancement by adaptive histogram equalization to improve the visibility of MCs with respect to the background, processing by multiscale wavelets and gray-level statistical techniques for feature extraction, clustering by the *k*-means algorithm for MC detection and, finally, using feature selection and a classifier based on a general regression neural network (GRNN) and multilayer perceptron (MLP) to classify MCs. Papadopoulos et al. [4] compared five image

\* Correspondence: joelq@ugto.mx

<sup>1</sup>Technical University of Madrid, 28040 Madrid, Spain

Full list of author information is available at the end of the article

enhancement algorithms for improving MC cluster detection in mammography. Halkiotis et al. [5] proposed mathematical morphology for MC extraction from a non-uniform background; in this scheme, a set of features is extracted from original mammograms to test two classifiers based on artificial neural networks, such as MLP, and a radial basis function (RBF) neural network. Fu et al. [6] proposed a method based on two stages. The purpose of the first stage is to locate the suspected MCs; this stage is based on mathematical morphology and border detection to segment the MCs. The second stage is based on feature extraction and selection from the MCs located in the first stage; in the final part of this latter stage, these features are used as an input vector to test two classifiers based on a GRNN and support vector machine (SVM).

In this paper, a method for detecting MCs in the regions of interest (ROIs) extracted from digitized mammograms is presented. The main purpose of this method is to provide an automatic MC detection system that can help radiologists to improve the diagnosis of breast cancer at an early stage. The method is based on image processing, pattern recognition and artificial intelligence techniques. The different stages of the method are as follows: image enhancement based on mathematical morphology operations, a novel *image sub-segmentation* approach based on possibilistic fuzzy *c*-means (PFCM) algorithm, which is compared with image segmentation by the *k*-means algorithm, feature extraction based on window-based features such as the mean and standard deviation and, finally, the use of a classifier based on an artificial neural network (ANN) to automatically detect MCs. Figure 1 shows a block diagram of the proposed method.

## 2 ROI image enhancement

Over the past several years, methodologies have been developed for the detection and/or classification of MCs, but the interpretation of MCs continues to be a

difficult task mainly because of their fuzzy nature, low contrast and low distinguishability from their surroundings. The difficulty of MC detection depends on factors such as size, shape and distribution with respect to MC morphology. Another important factor that also makes MC detection difficult is the fact that MCs are often located across non-homogeneous backgrounds, and owing to their low contrast against the background, their intensity may be similar to that of noise or other structures [7,8]. Therefore, in this paper, it is considered important to apply image enhancement.

Mathematical morphology is a discipline within the field of image processing that involves the structural analysis of images. The geometrical structure of an image is determined by locally comparing it with a pre-defined elementary set called a structuring element (*SE*). Image processing using morphological transformations is a process of information removal based on size and shape. In this process, irrelevant image content is selectively eliminated; thus, essential image features can be enhanced. Morphological operations are based on the relationships between two sets: an input image, *I*, and a processing operator, the *SE*, which is usually much smaller than the input image. By selecting the shape and size of a structuring element, different results can be obtained in the output image. The fundamental morphological operations are erosion and dilation.

The contrast can be defined as the difference in intensity between an image structure and its background. By combining morphological operations, several image processing tasks can be performed; however, in this work, we focus on those morphological operations that achieve contrast enhancement. In [8], a contrast enhancement technique using mathematical morphology is presented, called morphological contrast enhancement. Morphological contrast enhancement is based on morphological operations known as top-hat and bottom-hat transforms, which were proposed in [9]. A top-hat is a residual filter that preserves those features in an image that

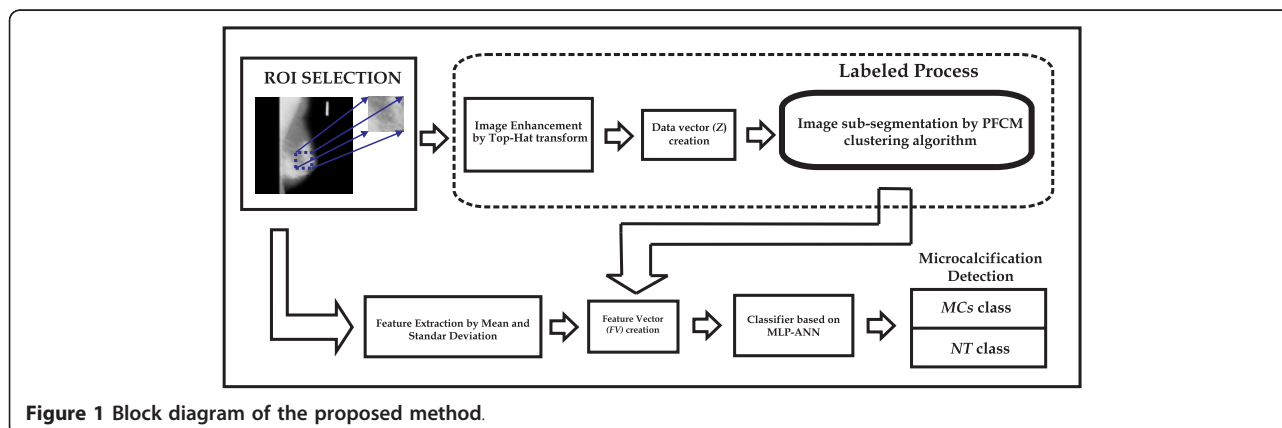


Figure 1 Block diagram of the proposed method.

can fit within 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:

$$I_T(x, y) = I(x, y) - [(I(x, y) \ominus SE) \otimes SE] \quad (1)$$

where  $I(x, y)$  is the input image,  $I_T(x, y)$  is the transformed image,  $SE$  is the structuring element,  $\ominus$  represents the morphological erosion operation,  $\oplus$  represents the morphological dilation operation, and  $-$  represents the image subtraction operation.  $[(I(x, y) \ominus SE) \oplus SE]$  is also known as the morphological opening operation. In previous works such as [8,10], this technique was used to obtain satisfactory results in MC detection.

### 3 Image segmentation by partitional clustering algorithms

Image segmentation is an important task in the field of image processing and computer vision and involves the identification of objects or regions with the same features in an image. The aim of image segmentation is to divide an image into non-overlapping subregions that are homogeneous with respect to some features such as gray-level intensity or texture. The level to which the subdivision is carried out depends on the problem being solved [11].

Depending on the specific application, several methods based on different principles have been used for image segmentation, such as histogram thresholding [12,13], edge detection [14,15], region growing [16-18], fractal models [19-22], ANNs [23], swarm-based algorithms [24] and clustering techniques [3,25-29].

In this paper, partitional clustering algorithms are considered for image segmentation, because of the great similarity between image segmentation and clustering, although clustering was developed for *feature space*, whereas segmentation was developed for the *spatial domain* of an image.

The clustering techniques represent non-supervised pattern classification into groups or classes. The partitional clustering techniques are based on cluster analysis, which is the organization of a set of patterns (vector of measurements or a point in a  $d$ -dimensional space) into clusters based on similarity [30]. In the context of image segmentation, the set of patterns can be represented by an image in a  $d$ -dimensional space that depends on the number of features used to represent the pixels, where each point in this  $d$ -dimensional space will be named a *pixel pattern*. Within the same context, the clusters correspond to some semantic meaning in the image, which is referred to as an object. Therefore,

the main goal of the clustering process is to obtain groups or classes from an unlabeled data set based on their similarities to facilitate further knowledge extraction. The similarity is evaluated according to a distance measure between the patterns and the prototypes or centers of the groups, and each pattern is assigned to the nearest or most similar prototype. However, this process must distribute all of the data to the different groups, even if some pixels are not very representative of the group as a whole [26]. In the field of medical imaging, segmentation plays an important role because it facilitates the delineation of anatomical structures and other regions that can be of interest. For the specific case of MC detection, several works based on image segmentation using partitional clustering algorithms have been proposed, such as [3,25-27]. Two clustering techniques based on partitional clustering algorithms are compared in this paper to improve the MC detection.

#### 3.1 $k$ -means

The  $k$ -means or hard  $c$ -means (HCM) algorithm [31] is one of the simplest unsupervised learning algorithms that can solve the well-known clustering problem. The objective of the clustering algorithms is to cluster a given data set into several groups such that the data within a group are more similar to one another than those outside the group. Achieving such a partition requires a similarity measure that considers two vectors and returns a value reflecting their similarity. The  $k$ -means algorithm partitions a given data set into  $c$  clusters and computes cluster centers  $V = [v_1, v_2, \dots, v_k]$ , so that the following objective function can be minimized.

$$J(Z; U, V) = \sum_{i=1}^c \sum_{k=1}^N \mu_{ik} \|z_k - v_i\|^2 \quad (2)$$

where  $\|z_k - v_i\|^2$  is the chosen distance measure between a data point  $z_k$  and the cluster  $v_i$  is an indicator of the distance of the data points from their cluster prototypes.  $V = [v_1, v_2, \dots, v_k]$  is the vector of prototypes of the  $c$  clusters, which are calculated according to:

$$v_i = \frac{1}{|A_i|} \sum_{z_k \in A_i} z_k \quad (3)$$

where  $|A_i|$  represents the number of data points belonging to cluster  $i$ .

To clarify, the procedures of the  $k$ -means algorithm are described as follows:

1. Initialize the cluster center  $v_i$ ,  $i = 1, \dots, c$ . This is typically achieved by randomly selecting  $c$  points from the data set.

2. Determine  $u_{ik}$ ,  $i = 1, 2, \dots, c$ ,  $k = 1, 2, \dots, N$ , by equation (4)

$$U = \mu_{ik} = \begin{cases} 1 & \text{if } \|z_k - v_i\|^2 \leq \|z_k - v_j\|^2, \quad \forall j \neq i. \\ 0 & \text{otherwise.} \end{cases} \quad (4)$$

3. Compute the objective function according to (2). Stop if either it has converged or the improvement is below a threshold.

4. Update the cluster center  $v_i$  using (3), and then proceed to Step 2.

### 3.2 PFCM clustering algorithm

The PFCM is one of the most recently developed partitioning clustering algorithms, which has the advantages of the fuzzy  $c$ -means (FCM) as well as the possibilistic  $c$ -means (PCM) algorithms. The FCM has a constraint that makes it very sensitive to outliers. To solve the problem of constraint of the FCM, Krisnapuram and Keller [32] developed the clustering algorithm PCM, which allows us to identify the *degree of typicality* that a data point has with respect to the group to which it belongs. The PCM has the problem, however, that sometimes the prototypes of clusters coincide, generating erroneous partitions of the feature space; for this reason, the PCM is not always successful. To solve the problems of the FCM (outlier sensitivity) and PCM (coincident clusters) clustering algorithms, Pal et al. [33] proposed a hybridized PFCM clustering model, where the function to be optimized is given by Equation 5:

$$J_{pfcM}(\mathbf{Z}; \mathbf{U}, \mathbf{T}, \mathbf{V}) = \sum_{i=1}^c \sum_{k=1}^N (a\mu_{ik}^m + bt_{ik}^\eta) \times \|z_k - v_i\|^2 + \sum_{i=1}^c \gamma_i \sum_{k=1}^N (1 - t_{ik})^\eta, \quad (5)$$

and is subject to the constraints  $\sum_{i=1}^c \mu_{ik} = 1 \forall k$ ;  $0 \leq \mu_{ik}, t_{ik} \leq 1$  with the constants  $a > 0$ ,  $b > 0$ ,  $m > 1$  and  $\eta > 1$ . The values of  $a$  and  $b$  represent the relative importance of membership and typicality values in the computation of the prototypes, respectively. The parameters  $m$  and  $\eta$  represent the absolute weight of the membership value and typicality value, respectively. To reduce the effect of outliers, one can set  $b > a$  and  $m > \eta$ .

*Theorem PFCM* [33]: If  $D_{ikA} = \|z_k - v_i\| > 0$ , for every  $i, k$ ,  $m > 1$ ,  $\eta > 1$ , and if

$Z$  contains at least  $c$  distinct data points, then  $(U, T, V) \in M_{fcm} \times M_{pcm} \times \mathfrak{R}^{c \times N}$  may minimize  $J_{pfcM}$  only if:

$$\mu_{ik} = \left( \sum_{j=1}^c \left( \frac{D_{ikA_i}}{D_{jkA_i}} \right)^{2/(m-1)} \right)^{-1} \quad (6)$$

$1 \leq i \leq c; \quad 1 \leq k \leq N$

$$t_{ik} = \frac{1}{1 + \left( \frac{b}{\gamma_i} D_{ikA_i}^2 \right)^{1/(\eta-1)}} \quad (7)$$

$1 \leq i \leq c; \quad 1 \leq k \leq N$

$$v_i = \sum_{k=1}^N (a\mu_{ik}^m + bt_{ik}^\eta) z_k / \sum_{k=1}^N (a\mu_{ik}^m + bt_{ik}^\eta), \quad (8)$$

$1 \leq i \leq c.$

$$\gamma_i = K \frac{\sum_{k=1}^N \mu_{ik}^m \|z^k - v_i\|^2}{\sum_{k=1}^N \mu_{ik}^m} \quad (9)$$

The iterative process of this algorithm is presented in [33].

To segment the MCs in ROI images, a novel technique based on the PFCM clustering algorithm is used. This technique is called *image sub-segmentation* and was proposed by Ojeda-Magaña et al. [26].

*Proposed approach for the detection of MCs by sub-segmentation*

1. Obtain the data vector.
2. Assign a value to the parameters  $(a, b, m, \eta)$ .
3. Segment the image by taking into account the number of more representative regions, which in this case is two: suspicious region with the presence of the MCs ( $S_1$ ) and normal tissue ( $S_2$ ); the  $S_2$  region is considered to be devoid of MCs.
4. Run the PFCM algorithm to obtain:
  - The membership matrix  $U$ .
  - The typicality matrix  $T$ .
5. Obtain the maximum typicality value for each pixel.

$$T_{\max} = \max_i [t_{ik}], \quad i = 1, \dots, c. \quad (10)$$

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

$$T_1 = T_{\max} \geq \alpha \quad (11)$$

containing the typical pixels of both regions (*Stypical*<sub>1</sub>) and (*Stypical*<sub>2</sub>), and the second matrix:

$$T_2 = T_{\max} < \alpha \quad (12)$$

containing the atypical pixels of both regions (*Satypical*<sub>1</sub>) and (*Satypical*<sub>2</sub>); in this case, the atypical pixels are of most interest, especially the *atypical* pixels of ( $S_1$ ).

8. From the labeled pixels  $z_k$  of the  $T_1$  sub-matrix, the following subregions can be generated:

$$T_1 = \text{Stypical}_{11}, \dots, \text{Stypical}_{i1}, \quad i = 1, \dots, c. \quad (13)$$

and from the  $T_2$  sub-matrix:

$$T_2 = \text{Satypical}_{1+i}, \dots, \text{Satypical}_{2i}, \quad i = 1, \dots, c. \quad (14)$$

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

$$S_i = \text{Stypical}_i \cup \text{Satypical}_{i+c}. \quad (15)$$

9. Select the sub-matrix  $T_1$  or  $T_2$  of interest for the corresponding analysis.

In this work,  $T_2$  is the sub-matrix of interest.

#### 4 Microcalcification classification by ANN

Artificial neural networks (ANNs) are biologically inspired networks based on the neuron organization and decision-making process of the human brain [34]. In other words, they are mathematical models 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 the 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. An ANN can approximate the function of multiple inputs and outputs. As a consequence, ANNs can be used for a variety of applications, among which are classification in medical applications [3,5,23,35], descriptive modeling, clustering, function approximation, time series prediction [36] and sonar or radar detection [37]. Classification is one of the most frequently encountered

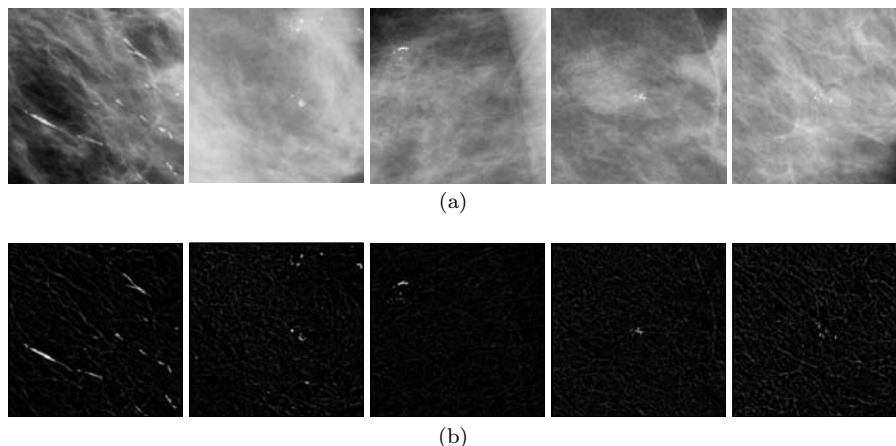
decision-making tasks in human activity. A classification problem occurs when an object needs to be assigned to a predefined group or class based on a number of observed patterns related to that object. In this paper, a classifier based on an ANN is used, with the aim of classifying patterns such as those that correspond to pixels belonging to healthy tissue or patterns that correspond to pixels belonging to microcalcifications, which we will call normal tissue class NT or MCs class, respectively. For this purpose, a multilayer perceptron (MLP) is used. The MLP is the most popular ANN for many practical applications, such as pattern recognition applications. The functionality of the MLP topology is determined by a learning algorithm, the back propagation (BP) [38], which is based on the method of steepest descent [39]. In upgrading connection weights, it is the algorithm most commonly used by the ANN scientific community.

#### 5 Methodology and results

To test our method, a set of ten ROI images were selected from several mammograms of the mini-MIAS database provided by the Mammographic Image Analysis Society (MIAS) [40]. The size of each mammogram from this database is  $1,024 \times 1,024$  pixels, with a spatial resolution of  $200\mu\text{m}/\text{pixel}$ . These mammograms were reviewed by an expert radiologist, and all abnormalities were identified and classified. The areas in which abnormalities such as MCs were located were taken as ROIs. In this work, ROI images measuring  $256 \times 256$  pixels were used. Figure 2 a shows some ROI images used in this work.

##### 5.1 Morphological enhancement

The morphological top-hat transform is used to enhance ROI images, with the aim of detecting objects that differ in brightness from the surrounding background; in our



**Figure 2** a Original ROI images. b ROI images processed by the top-hat transform.

case, it was used to increase the contrast between the MCs and the background. During image enhancement, the same  $SE$  at different sizes,  $3 \times 3$ ,  $5 \times 5$ ,  $7 \times 7$ , was applied to perform the top-hat transform. The  $SE$  used in this work was a flat disk-shaped  $SE$ . Figure 2 shows the original ROI images processed by the top-hat transform with a  $SE$  of size  $7 \times 7$ .

## 5.2 Image segmentation by clustering

### 5.2.1 Data vector creation

A *data vector*  $Z$  for each ROI is generated for each of the images obtained from the previous stage. Thus, a unidimensional vector ( $\mathbf{x}_{se}$ ) is built by mapping the images to the pixels as follows:

$$\{[I_T(x, \gamma)]_{1 \leq x \leq R, 1 \leq \gamma \leq C}\}_{se} \rightarrow \mathbf{x}_{se} = \{x_{se}^{(q)}\}_{q=1, \dots, R \times C} \quad (16)$$

where  $se$  is the size of the  $SE$ ,  $x_{se}^{(q)}$  is the gray-level of the  $q$ th pixel of  $I_T$  when the image is decomposed row by row, and  $R$  and  $C$  correspond to the size of the image. Then, the data vector  $Z$  can be written as follows:

$$Z = [\mathbf{x}_{3 \times 3}, \mathbf{x}_{5 \times 5}, \mathbf{x}_{7 \times 7}]^T \quad (17)$$

For *data vector*  $Z$ , two proposed clustering techniques are then applied to obtain a label for each pattern belonging to each cluster of the partition of feature space, where only one cluster corresponds to MCs,

which generally appear in a group of just a few patterns (pixels), and the remaining clusters correspond to normal (healthy) tissue.

The initial conditions and results for each proposed clustering technique are presented below.

### 5.2.2 Segmentation by $k$ -means

The initial conditions for this approach are as follows:

- Cluster number: 2 to 4.
- Prototypes: initialized as random values.
- Distance measure: Euclidean distance function.

Figure 3 shows segmented ROI images with different cluster values obtained after applying the proposed  $k$ -means algorithm to the *data vector*  $Z$ .

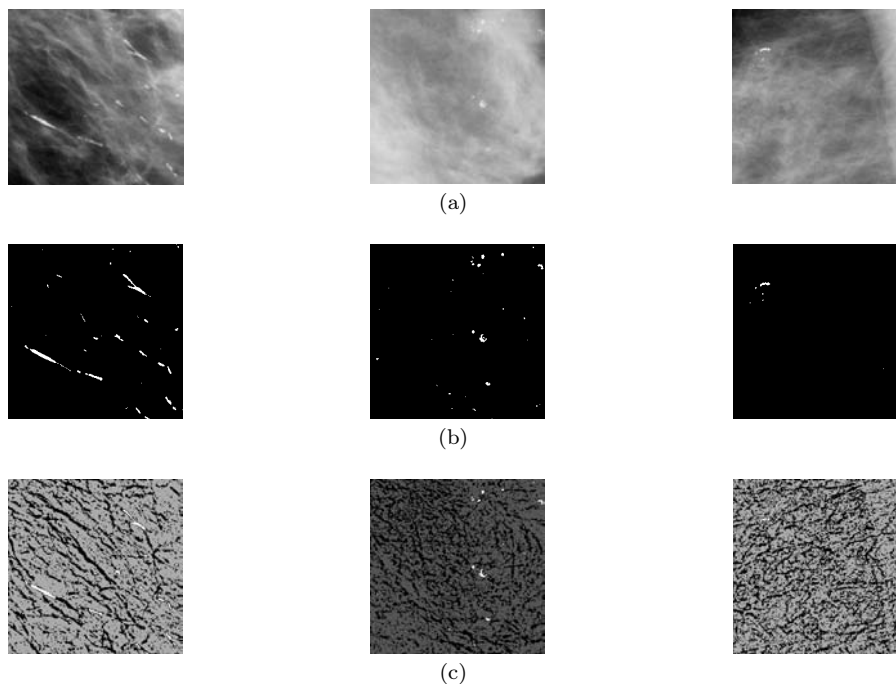
### 5.2.3 Sub-segmentation by PFCM

In this case, the approach presented in Section 3.2 is applied, and the initial conditions are as follows:

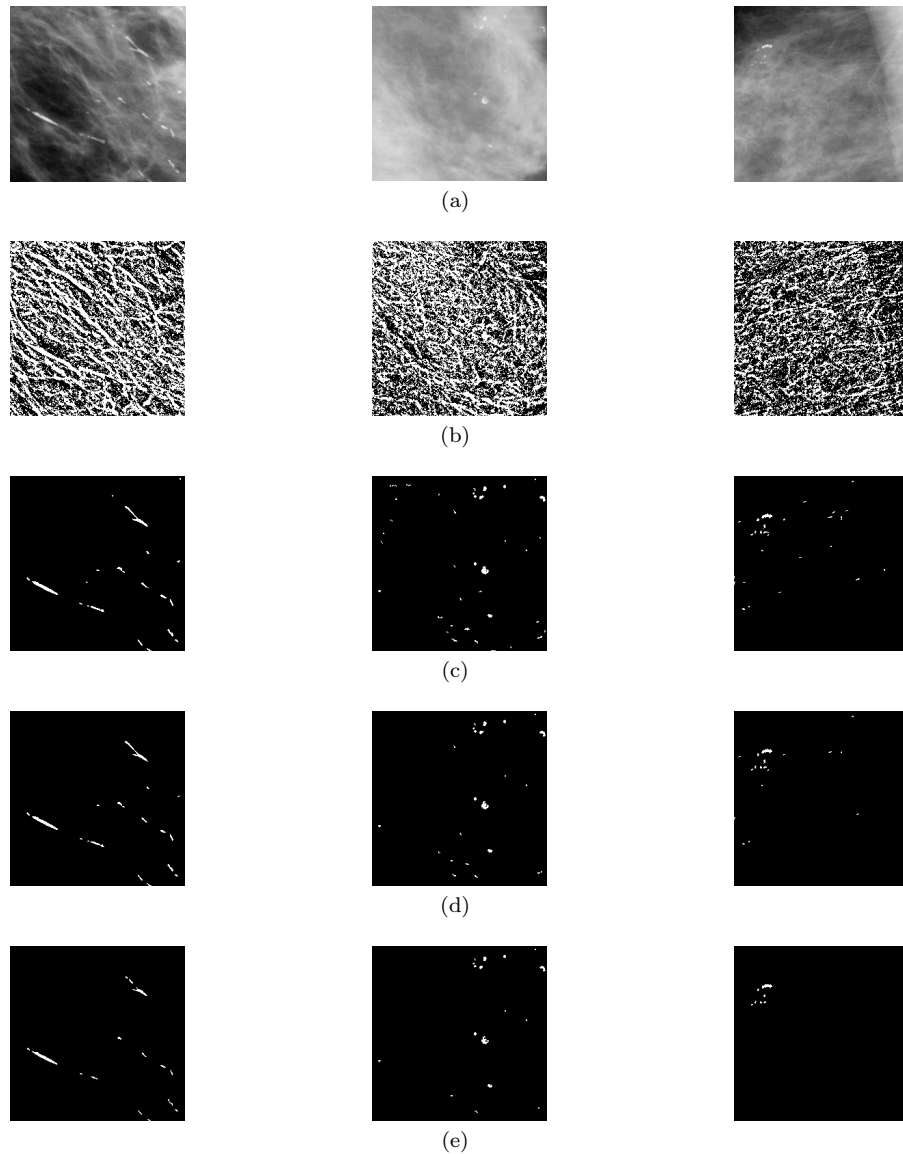
- Cluster number: 2.
- Prototypes: initialized as random values.
- Distance measure: Euclidean distance function.
- $a = 1$ ,  $b = 2$ ,  $m = 2$ ,  $\eta = 2$ ,  $\alpha = 0.04$ ,  $\alpha = 0.03$ ,  $\alpha = 0.02$ .

Figure 4 shows segmented ROI images with different threshold values ( $\alpha$ ) obtained after applying the approach presented in Section 3.2 to the *data vector*  $Z$ .

According to the results obtained from the clustering process by  $k$ -means and PFCM, Table 1 shows the number of patterns assigned to classes MCs and NT, respectively, for our set of ten ROI images.



**Figure 3** Image segmentation by  $k$ -means. **a** Original ROI images. **b** The results obtained from the 2nd partition. **c** The results obtained from the 3rd partition.



**Figure 4** **a** Original ROI images. **b** Segmentation of the ROIs using the membership matrix  $U$  into 2 groups. Final image sub-segmentation by PFCM using the typicality matrix  $T$  and threshold values of  $\alpha =$  (c) 0.04, (d) 0.03, (e) 0.02.

#### 5.2.4 Feature extraction

Two window-based features, such as the mean and standard deviation defined in Equations 18 and 19, respectively, are extracted.

$$I_{\mu}(x, y) = \frac{1}{R \times C} \sum_{x=1}^R \sum_{y=1}^C f(x, y) \quad (18)$$

$$I_{\sigma}(x, y) = \left[ \frac{1}{R \times C} \sum_{x=1}^R \sum_{y=1}^C (f(x, y) - I_{\mu}(x, y))^2 \right]^{1/2} \quad (19)$$

where  $I_{\mu}$ ,  $I_{\sigma}$  and  $f(x, y)$  represent the mean, standard deviation and the gray-level value of a pixel located in  $(x, y)$ , respectively. These features are extracted from original ROI images within rectangular windows; in this work, we used three different pixel block windows with sizes ( $ws$ ),  $3 \times 3$ ,  $5 \times 5$  and  $7 \times 7$ . In our work, each image obtained by this process is considered a feature that can be used to generate a set of patterns. In this set of patterns, there are patterns that represent the MCs and NT classes. We refer to this set of patterns as the feature vector (FV). We know a priori that, for each image used in this work, there are pixels belonging to

**Table 1 Number of patterns assigned to MCs and NT**

Class	Number of patterns by <i>k</i> -means	Number of patterns by sub-segmentation with PFCM
NT	652,965	652,716
MCs	2,395	2,644

the MCs and NT class. This FV is considered an input vector for the classifier. The FV is formed as follows:

$$FV = [\mathbf{i}_{\mu 3 \times 3}, \mathbf{i}_{\sigma 3 \times 3}, \mathbf{i}_{\mu 5 \times 5}, \mathbf{i}_{\sigma 5 \times 5}, \mathbf{i}_{\mu 7 \times 7}, \mathbf{i}_{\sigma 7 \times 7}] \quad (20)$$

where:

$$[\{I_T(x, \gamma)\}_{1 \leq x \leq R, 1 \leq \gamma \leq C}]_{us} \rightarrow \mathbf{i}_{\mu_{us}} = \left\{ i_{\mu_{us}}^{(q)} \right\}_{q=1, \dots, R \times C} \quad (21)$$

$$[\{I_\sigma(x, \gamma)\}_{1 \leq x \leq R, 1 \leq \gamma \leq C}]_{us} \rightarrow \mathbf{i}_{\sigma_{us}} = \left\{ i_{\sigma_{us}}^{(q)} \right\}_{q=1, \dots, R \times C} \quad (22)$$

The labels of the two classes of the FV were obtained by the previous process. Owing to the large number of patterns that do not belong to the MCs class, with respect to the number of patterns that do belong to the MCs class, balancing was performed. Table 2 shows the subsets of the patterns for the MCs and NT classes.

### 5.3 Microcalcification classification by ANN

A MLP was used to classify the patterns as NT or MCs, with the purpose of automatically identifying MCs in ROIs extracted from mammograms. To comparatively evaluate the performance of the classifiers, in this particular case, different network structures were trained and tested with the same training data set and the same testing data set. The best obtained results possessed the following structure and parameters:

1. Number of input neurons equal to the number of attributes in FV: 6.
2. Number of hidden layers: 1.
3. Hidden neurons: see Table 4).
4. Output neurons: 1 (all classifications present two classes).
5. Learning rate: 1.
6. Activation function is sigmoidal with values between [0,1].

**Table 2 Results of balancing**

Class	Number of patterns by <i>k</i> -means	Number of patterns by sub-segmentation with PFCM
NT	23,950	26,440
MCs	2,395	2,644

**Table 3 Number of patterns used for training and testing for each classifier**

		Numbers of Sample			
		FV <sub>MCs/NT</sub>	Training	Testing	Total
<i>k</i> -means	MCs		1,912	483	2,395
	NT		19,164	4,786	23,950
PFCM	MCs		2,097	547	2,644
	NT		21,171	5,269	26,440

7. All weights randomly initialized.
8. Training phase: back propagation (BP).
9. Test training conditions:
  - (a) epochs: 2000.
  - (b) mean squared error (MSE): 0.001.

In this paper, we used patterns extracted from the FV set to train and test our classifiers: 80% of the patterns were used for training, and 20% of the patterns were used for testing (see Table 3).

Table 4 shows the optimal network structure and parameters for each FV.

A confusion matrix to determine the probability of MC detection versus the probability of false MC detection was built. Table 5 shows the performance of the classifiers presented in this work. The performance of the proposed method was evaluated by means of ROC (receiver operating characteristics) curve analysis. The ROC curve is a two-dimensional measure of classification performance and is widely used in biomedical applications to assess the performance of diagnostic tests. The ROC curve is a plot of the *sensitivity* versus *specificity* for the different possible cut-points of a diagnostic test. Figure 5 shows the ROC curve and the area under the curve (*AUC*) for the classifiers with different network structures used in this work.

Finally, Figure 6 shows the results of MC detection in the ROIs using the methodology proposed in this paper.

**Table 4 The best network structure and parameters for each database**

Data set (FV)	Network structure			MSE
	I	HL	O	
<i>k</i> -means	6	15	1	0.001
PFCM	6	12	1	0.001



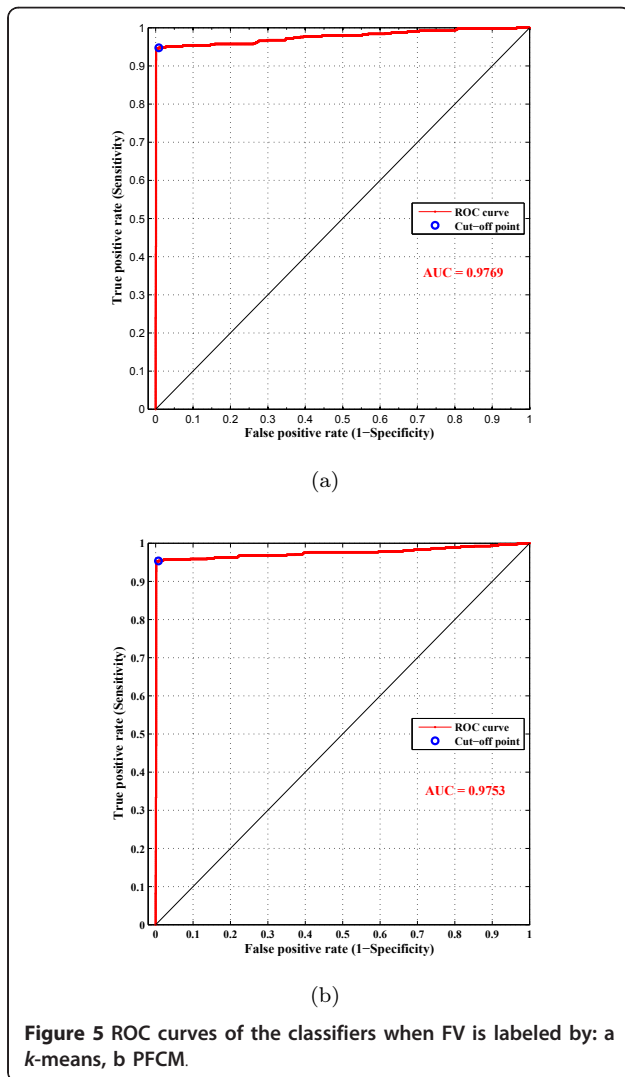
**Table 5 Confusion matrices and performance of the classifiers**

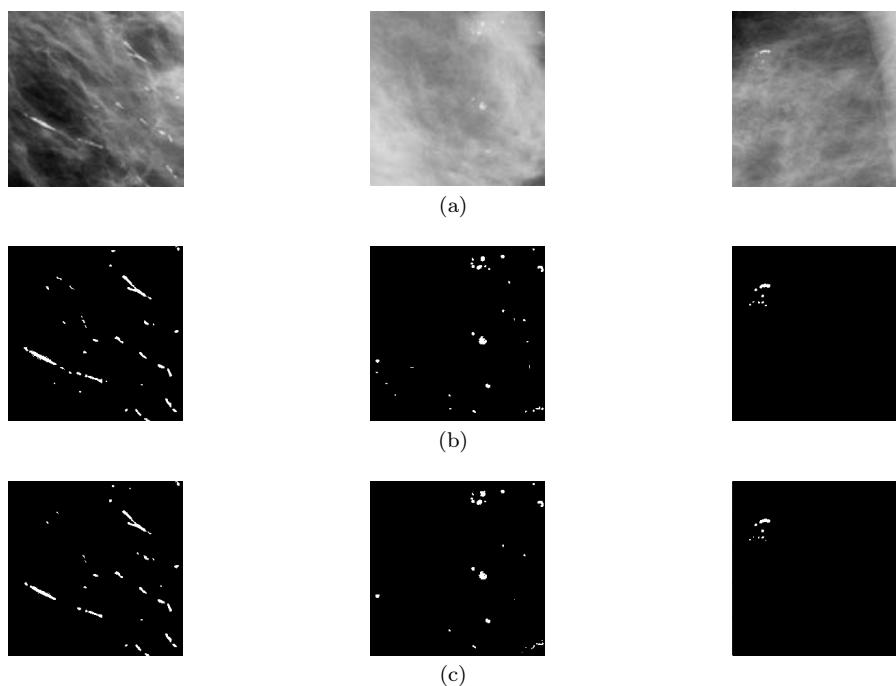
Classifier MLP-BP	Desired results	Output		Sensitivity (%)	Specificity (%)	Total class. Accuracy (%)
		MC	NT			
<i>k</i> -means						
Structure <sub>1</sub>	MC	475	8	98.34	99.44	99.31
6 : 15 : 1	NT	27	4,759			
PFCM						
Structure <sub>2</sub>	MC	534	13	97.62	99.49	99.31
6 : 12 : 1	NT	27	5,242			

**6 Discussion and conclusions**

According to the performance of the classifiers as determined by means of the ROC curves (Figure 5) and the final images obtained (Figure 6), the proposed method is a promising alternative for automatically detecting MCs

in ROIs extracted from digitized mammograms. This method involves several techniques that contribute to the MCs detection stage. The image segmentation stage is one of the most difficult stage when using partitional clustering algorithms, because these clustering algorithms are applied in the features space. Therefore, if the image contains noise or is not very homogeneous, image segmentation by clustering can be inaccurate. Thus, an image processing technique based on mathematical morphology was used to solve this problem. In the segmentation stage, two partitional clustering algorithms were used: *k*-means and PFCM. The *k*-means is the most popular technique, and its advantages and drawbacks are well known. With the PFCM, a new method for image segmentation called *image sub-segmentation* was used, in which the degrees of typicality of each data point were used to partition an image into two regions: one region with tissue suspected of harboring MCs and the other with normal (healthy) tissue. Then, the most atypical data points (pixels) of each region were identified; these data include possible abnormalities present in these regions, especially the region suspected of possessing MCs, because these atypical data, or abnormalities, represent the pixels belonging to potential MCs. For the ROI images used in this paper, both clustering algorithms used to perform image segmentation gave good results, although these results depend largely on good feature extraction and, in this paper, on the image enhancement stage. Once the MCs were detected from the original ROIs, window-based features such as the mean and standard deviation were extracted, which were then used as input vectors in a classifier. To perform this classification task, ANNs proved to be an excellent alternative. In this paper, a classifier based on the MLP was used. In the ROI images, the MCs class represented a lower percentage of pixels with respect to the number of pixels belonging to the healthy or normal tissue class. Therefore, balancing between patterns belonging to the MCs class and to the NT class was performed to obtain better results during the classification stage. Finally, according to the results





**Figure 6** a Original ROI images. MCs detection using: b image segmentation by  $k$ -means and the classifier with structure<sub>1</sub>, c image sub-segmentation by PFCM and the classifier with structure<sub>2</sub>.

obtained by applying our proposed method to these ROI images, the implemented method can detect pixels corresponding to microcalcifications or healthy tissue, thus fulfilling the aim of this paper.

### 7 Acknowledgements

The authors wish to thank the Group for Automation in Signal and Communications (GASC) of the Technical University of Madrid, the Laboratorio de Inteligencia Computacional (LABINCO) of the Guanajuato University, The National Council for Science and Technology (CONACYT), the Department of Project Engineering (CUCEI) of the University of Guadalajara and the Ph.D. Bernhard Angele.

### Author details

<sup>1</sup>Technical University of Madrid, 28040 Madrid, Spain <sup>2</sup>University of Guadalajara, 45101 Zapopan Jalisco, Mexico <sup>3</sup>University of Guanajuato, 36885 Salamanca Guanajuato, Mexico

### Competing interests

The authors declare that they have no competing interests.

Received: 15 May 2011 Accepted: 24 October 2011

Published: 24 October 2011

### References

1. N Pal, B Bhowmick, S Patel, S Pal, J Das, A multi-stage neural network aided system for detection of microcalcifications in digitized mammograms. *Neurocomputing*. **71**(13-15), 2625–2634 (2008). doi:10.1016/j.neucom.2007.06.015
2. L Wei, Y Yang, R Nishikawa, Microcalcification classification assisted by content-based image retrieval for breast cancer diagnosis. *Pattern Recognit.* **42**(6), 1126–1132 (2009). doi:10.1016/j.patcog.2008.08.028
3. A Vega-Corona, A Álvarez, D Andina, Feature Vectors Generation for Detection of Microcalcifications in Digitized Mammography Using Neural Networks, vol. 2687. *Artificial Neural Nets Problem Solving Methods*, LNCS, 583–590 (2003)
4. A Papadopoulos, D Fotiadis, L Costaridou, Improvement of microcalcification cluster detection in mammography utilizing image enhancement techniques. *Comput Biol Med.* **38**(10), 1045–1055 (2008). doi:10.1016/j.compbiomed.2008.07.006
5. S Halkiotis, T Botsis, M Rangoussi, Automatic detection of clustered microcalcifications in digital mammograms using mathematical morphology and neural networks. *Signal Proces.* **87**(7), 1559–1568 (2007). doi:10.1016/j.sigpro.2007.01.004
6. J Fu, S Lee, S Wong, J Yeh, A Wang, H Wu, Image segmentation feature selection and pattern classification for mammographic microcalcifications. *Comput Med Imaging Graph.* **29**(6), 419–429 (2005)
7. H Cheng, X Cai, X Chen, L Hu, X Lou, Computer-aided detection and classification of microcalcifications in mammograms: a survey. *Pattern Recognit.* **36**(12), 2967–2991 (2003). doi:10.1016/S0031-3203(03)00192-4
8. M Wirth, M Fraschini, J Lyon, Contrast enhancement of microcalcifications in mammograms using morphological enhancement and non-flat structuring elements, in *17th IEEE Symposium on Computer-Based Medical System*, pp. 134–139 (2004)
9. F Meyer, Iterative image transformations for an automatic screening of cervical smears. *J Histochem Cytochem.* **27**(1), 128–135 (1979). doi:10.1177/27.1.438499
10. T Stojić, B Reljin, Enhancement of microcalcifications in digitized mammograms: Multifractal and mathematical morphology approach. *FME Trans.* **38**, 1–9 (2010)
11. R Gonzalez, R Woods, *Digital Image Processing* (Prentice Hall, 2002)
12. J Ge, M Hadjiiski, B Sahiner, J Wei, M Helvie, C Zhou, H Chan, Computer-aided detection system for clustered microcalcifications: comparison of performance on full-field digital mammograms and digitized screen-film mammograms. *Phys Med Biol.* **52**(4), 981–1000 (2007). doi:10.1088/0031-9155/52/4/008
13. Y Wu, Q Huang, Y Peng, W Situ, Detection of microcalcifications in digital mammograms based on dual-threshold, in *International Workshop on Digital Mammography, IWDM*, LNCS, **4046**, 347–354 (2006). doi:10.1007/11783237\_47

14. G Veni, E Regentova, L Zhang, Detection of Clustered Microcalcifications with Susan Edge Detector, Adaptive Contrast Thresholding and Spatial Filters. Lecture notes in computer science. **5112**, 837–843 (2008). doi:10.1007/978-3-540-69812-8\_83
15. A Jevtic, J Quintanilla-Domínguez, M Cortina-Januchs, D Andina, Edge detection using ant colony search algorithm and multiscale contrast enhancement, in *Systems, Man and Cybernetics, SMC 2009. IEEE International Conference on*, 2193–2198 (2009)
16. G Qiu, Z Jianhua, C Shengyong, A Todd-Pokropek, Automatic segmentation of micro-calcification based on sift in mammograms, in *International Conference on BioMedical Engineering and Informatics*. **2**, 13–17 (2008)
17. I Bankman, T Nizialek, I Simon, O Gatewood, I Weinberg, W Brody, Segmentation algorithms for detecting microcalcifications in mammograms. *IEEE Trans Inf Technol Biomed*. **1**(2), 141–149 (1997). doi:10.1109/4233.640656
18. A Rojas-Domínguez, AK Nandi, Toward breast cancer diagnosis based on automated segmentation of masses in mammograms. *Pattern Recognit*. **42**(6), 1138–1148 (2009). doi:10.1016/j.patcog.2008.08.006
19. T Stojić, I Reljin, B Reljin, Adaptation of multifractal analysis to segmentation of microcalcifications in digital mammograms. *Phys A Stat Mech Appl*. **367**, 494–508 (2006)
20. JA Piuella, D Andina, K McInnes, AM Tarquis, Wavelet analysis in a structured clay soil using 2-d image. *Nonlinear Proces Geophys*. **14**, 425–234 (2007). doi:10.5194/npg-14-425-2007
21. C Dar-Ren, C Ruey-Feng, C Chii-Jen, H Minng-Feng, K Shou-Jen, CS-T Che, H Shin-Jer, WK Moon, Classification of breast ultrasound image using fractal features. *Clin Imaging*. **29**(4), 235–245 (2005). doi:10.1016/j.clinimag.2004.11.024
22. H Li, R Liu, S Lo, Fractal modeling and segmentation for the enhancement of microcalcifications in digital mammograms. *IEEE Trans Med Imaging*. **16**(6), 785–798 (1997)
23. B Verma, Impact of multiple clusters on neural classification of rois in digital mammograms, in *International Joint Conference on Neural Networks*, pp. 3220–3223 (2009)
24. A Jevtic, J Quintanilla-Domínguez, J Barron-Adame, D Andina, Image segmentation using ant system-based clustering algorithm, in *Soft Computing Models in Industrial and Environmental Applications, 6th International Conference SOCO 2011*, **87** (2011)
25. M Bhattacharya, A Das, Fuzzy logic based segmentation of microcalcification in breast using digital mammograms considering multiresolution. *Int Mach Vis Image Process Conf*, 98–105 (2007)
26. B Ojeda-Magaña, J Quintanilla-Domínguez, R Ruelas, D Andina, Images sub-segmentation with the pfc clustering algorithm, in *7th IEEE International Conference on Industrial Informatics*, pp. 499–503 (2009)
27. P Bougioukos, D Glotsos, S Kostopoulos, A Daskalakis, I Kalatzis, N Dimitropoulos, G Nikiforidis, D Cavouras, Fuzzy c-means-driven fnce contextual segmentation method for mammographic microcalcification detection. *Imaging Sci J*. **58**(3), 146–154 (2009)
28. G Yang, G Zhou, Y Yin, X Yang, K-means based fingerprint segmentation with sensor interoperability. *EURASIP J Adv Signal Process*. **2010** (2010)
29. J Mu, X Liu, M Kamocka, Z Xu, M Alber, E Rosen, D Chen, Segmentation, reconstruction, and analysis of blood thrombus formation in 3d 2-photonmicroscopy images. *EURASIP J Adv Signal Process*. **2010** (2010)
30. A Jain, M Murty, P Flynn, Data clustering: a review. *ACM Comput Surv*. **31** (1999)
31. J MacQueen, Some methods for classification and analysis of multivariate observations, in *Fifth Berkeley Symposium on Mathematical Statistics and Probability*. **1**, 281–297 (1967)
32. R Krishnapuram, JM Keller, A possibilistic approach to clustering. *IEEE Trans Fuzzy Syst*. **1**(2), 98–110 (1993). doi:10.1109/91.227387
33. N Pal, S Pal, J Keller, J Bezdek, A possibilistic fuzzy c-means clustering algorithm. *IEEE Trans Fuzzy Syst*. **13**(4), 517–530 (2005)
34. D Andina, D Pham, *Computational Intelligence for Engineering and Manufacturing*, 1st edn. (Springer, 2007)
35. A Papadopoulos, D Fotiadis, A Likas, An automatic microcalcification detection system based on a hybrid neural network classifier. *Artif Intell Med*. **25**(2), 149–167 (2002). doi:10.1016/S0933-3657(02)00013-1
36. M Cortina-Januchs, J Barrón-Adame, A Vega-Corona, D Andina, Prevision of industrial SO<sub>2</sub> pollutant concentration applying anns, in *7th IEEE International Conference on Industrial Informatics, INDIN*, pp. 510–515 (2009)
37. D Andina, J Sanz-Gonzalez, On the problem of binary detection with neural networks, in *Circuits and Systems Proceedings, Proceedings of the 38th Midwest Symposium on*. **1**, 554–557 (1995)
38. I Basheer, M Hajmeer, Artificial neural networks: fundamentals, computing, design, and application. *J Microbiol Methods*. **43**(1), 3–31 (2000). doi:10.1016/S0167-7012(00)00201-3
39. M Hagan, H Demuth, M Beale, *Neural Network Design* (PWS Pub Co, Boston, 1996)
40. J Suckling, J Parker, D Dance, The mammographic image analysis society digital mammogram database, in *Exerpta Medica International Congress Series*. **1069**, 375–378 (1994)

doi:10.1186/1687-6180-2011-91

**Cite this article as:** Quintanilla-Domínguez et al.: Improvement for detection of microcalcifications through clustering algorithms and artificial neural networks. *EURASIP Journal on Advances in Signal Processing* 2011 **2011**:91.

**Submit your manuscript to a SpringerOpen® journal and benefit from:**

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](http://springeropen.com)