

# Analysis of Mammograms Using Texture Segmentation

Joel Quintanilla-Domínguez<sup>1</sup>, Jose Miguel Barrón-Adame<sup>1</sup>,  
Jose Antonio Gordillo-Sosa<sup>1</sup>, Jose Merced Lozano-Garcia<sup>2</sup>, Hector Estrada-García<sup>2</sup>,  
Rafael Guzmán-Cabrera<sup>2</sup>

<sup>1</sup> Universidad Tecnológica del Suroeste del Estado de Guanajuato, UTSOE, Mexico

<sup>2</sup> Universidad de Guanajuato, Campus Irapuato-Salamanca, Mexico

guzmanc@ugto.mx

**Abstract.** Breast cancer is one of the most dangerous types of cancer among women around the world. It is also one of the leading causes of mortality in middle and old aged women. The World Health Organization's International Agency for Research on Cancer estimates that more than 1 million cases of breast cancer will occur worldwide annually, with 580,000 cases occurring in developed countries and the remainder in developing countries. In this paper, we present an effective methodology in order to detect clusters of Micro calcifications in digitized mammograms, based on the synergy of Image Processing, Pattern Recognition and Artificial Intelligence. The results obtained allow us to see the effectiveness of the proposed method.

**Keywords.** Segmentation, cancer detection, image processing.

## 1 Introduction

The risk of a woman developing breast cancer during her life time is approximately 11% [1]. The early detection of breast cancer is of vital importance for the success of treatment, with the main goal to increase the probability of survival for patients. Currently the most reliable and practical method for early detection and screening of breast cancer is mammography. Mammography is a technique used to visualize normal and abnormal structures within the breasts. Mammography is a special type of X-ray imaging used to create detailed images of the breast (Mammograms). However, achieving this early cancer detection is not an easy task. Micro calcifications (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, their variation in shape, orientation, brightness and diameter size [2]. But it is still a hard task to detect all the MCs in mammograms, because of the poor contrast with the tissue that surrounds them.

However, many techniques have been proposed to detect the presence of MCs in mammograms: image enhancement techniques, Artificial Neural Networks (ANN), wavelet analysis, Support Vector Machines (SVM), mathematical morphology, image analysis models, fuzzy logic techniques, etc. Image enhancement algorithms have been utilized for the improvement of contrast features and the suppression of noise. In [3] proposed five image enhancement algorithms for the detection of MCs in mammograms.

Bhattacharya and Das [4] proposed a method based on discrete wavelet transform due to its multiresolution properties with the goal to segment MCs in digital mammograms. Morphological Top-Hat algorithm was applied for contrast enhancement of the MCs. Fuzzy C-Means clustering (FCM) algorithm was implemented for intensity-based segmentation. Sung et al. [5] proposed an approach by means of mathematical morphology operations and wavelet transform to locate the MCs in digital mammogram.

In [6] proposed an algorithm that was tested over several images taken from the digital database for screening mammography for cancer research and diagnosis, and it was found to be absolutely suitable to distinguish masses and microcalcifications from the background tissue using morphological operators and then extract them through machine learning techniques and a clustering algorithm for intensity-based segmentation. Segmentation processes for the detection of textures, ROIs, lesions, tumors have also been used on photo-acoustic images [7] and thermographic images [8].

The remaining sections of this work are organized as follows: Section 2, presents the details of the proposed method. Section 3 presents the details of the proposed method and experimental results while the conclusions are presented in sections 4, respectively.

## **2 Methodology**

The mammograms used to train and test in this work were extracted from a mini-mammographic database provided by Mammographic Image Analysis Society (MIAS) [9]. This database contains 322 mammograms, 118 mammograms contain some abnormality, 66 are benign and 52 are malignant, and the remainder of the mammograms are diagnosed as normal.

The abnormalities found in these mammograms are MCs, Well-defined/circumscribed masses, ill-defined masses, speculated masses, architectural distortions and asymmetries. Each mammogram from the database is  $1024 \times 1024$  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. The place where these abnormalities such as MCs, have been located is known as, Region of Interest (ROI). The ROI images size in this work is  $256 \times 256$ .

The proposed approach is applied to each of the ROI images individually in order to show the obtained results by means of a segmented image. Several ROI images

from mammograms with dense tissue and the presence of MCs were selected to train and test the proposed approach.

Next, the morphological Top-Hat transform is used in order to enhance the ROI image, with the goal of detecting objects that differ in brightness from the surrounding background, in this case is to increase the contrast between the MCs clusters and background. Then, the same structuring element at with same size is applied. The size of structuring element used is  $3 \times 3$ . Fig. 1 shows original ROI images processed by Top-Hat transform. In the next stage two window-based features such as, mean and standard deviation were applied. They are extracted from enhancement images within a rectangular window, in this work a  $5 \times 5$ -pixel block window was used.

To carry out the grouping and labeling, we use two methods well known in the state of the art: k-means and SMO, these methods are briefly described below:

K-means is one of the simplest unsupervised learning algorithms that solve the well-known clustering problem. The procedure follows a simple and easy way to classify a given data set  $Z$  in a  $d$ -dimensional space, through a certain number of clusters (assume  $k$  clusters) fixed a priori. The main idea is to define  $k$  prototypes, one for each cluster. The next step is to take each point belonging to a given  $Z$  and associate it to the nearest prototype. When no point is pending, the first step is completed and an early group is done.

At this point it is necessary to re-calculate  $k$  new prototypes as centers of the clusters resulting from the previous step. After obtaining these  $k$  new prototypes, a new binding has to be done between the same data set points and the nearest new prototype. A loop has been generated. As a result of this loop we may notice that the  $k$  prototypes change their location step by step until no more changes are done. In other words, prototypes do not move any more.

The initial conditions used in this work for this method were:

- Cluster number takes values 2 to 6.
- Prototypes were initialized as random values.
- Euclidean distance function.
- Maximum iteration number: 100.

Self-organizing maps (SMO), are simple analogues to the brain's way to organize information in a logical manner. The main purpose of this neural information processing is the transformation of a feature vector of arbitrary dimension drawn from the given feature space into simplified generally two-dimensional discrete maps. A SOM network performs the transformation adaptively in a topological ordered fashion. This type of neural network utilizes an unsupervised learning method, known as competitive learning, and is useful for analyzing data with unknown relationships. The basic SOM Neural Network consists of the input layer, and the output (Kohonen) layer which is fully connected with the input layer by the adjusted weights (prototype vectors). The number of units in the input layer corresponds to the dimension of the data.

There are several steps in the application of the algorithm. These are competition and learning, to get the winner in the process. In the training (learning) phase, the

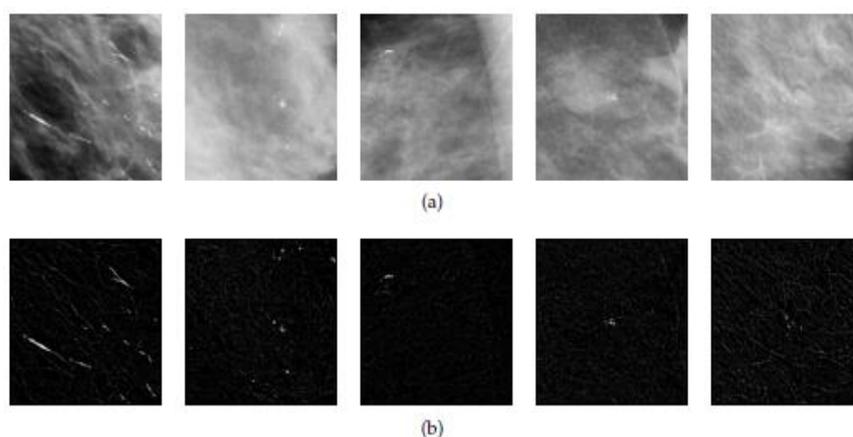
SOM forms an elastic net that folds onto the “cloud” formed by the input data. Similar input vectors should be mapped close together on nearby neurons, and group them into clusters. If a single neuron in the Kohonen layer is excited by some stimulus, neurons in the surrounding area are also excited. That means for the given task of interpreting multidimensional image data, each feature vector  $x$ , which is presented to the four neurons of the input layer, typically causes a localized region of active neurons against the quiet background in the Kohonen layer.

The initial conditions for this method were:

- Network structure [4 k], k takes values 2 to 16.
- Weight vector was initialized as random values.
- Topology function: Hexagonal layer.
- Distance function: Euclidean distance.
- Maximum epoch: 100.

### 3 Results

In this work, each image obtained after applying the image enhancement process as well as the images obtained by window-based features, are considered as a feature to generate a set of patterns that represent the MCs and the normal tissue. Each pattern is constructed from gray level intensity of pixels of the obtained images, each pattern generated is called features vector, where these features vector represents a point in d-dimensional space. It is known each of the images used there are pixels belonging to MCs and to normal tissue, then each analyzed pattern belong to one of two possible classes, i.e. there are patterns that belong to the set  $Q_1$  if correspond to MCs and patterns belonging to normal tissue  $Q_0$ .



**Fig. 1.** (a)Original ROI images. (b) ROIs images processed by Top-Hat transform.

In Table 1 shows the number of patterns assigned to classes  $Q_0$  and  $Q_1$  obtained.

**Table 1.** Number of patterns assigned to  $Q_1$  and  $Q_0$ .

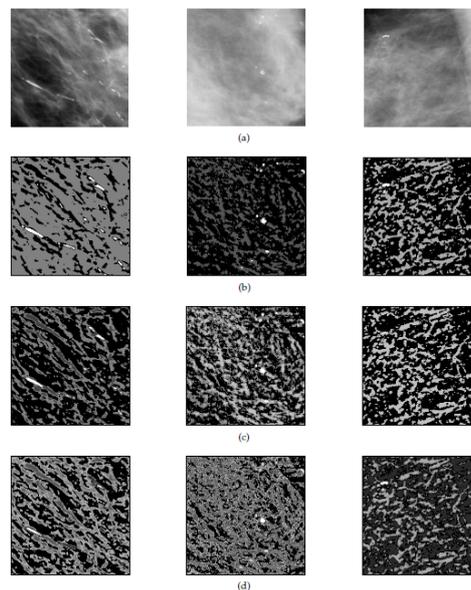
Label	Number of patterns by k-means	Number of patterns by SOM
$Q_0$	588181	583081
$Q_1$	1643	6743

Due to the large number of patterns of the class that do not belong to MCs with respect to the number of patterns that belong to the class of MCs a balancing was performed, see Table 2, The network parameters such as network size and architecture (number of nodes, hidden layers etc.), and gain parameters were kept the same. For all cases the neural network had one hidden layer with eight hidden nodes. In order to determine the network structure and metaplasticity parameters, the same network parameters applied in [10] and [11] were used.

**Table 2.** Results of balancing.

Label	Number of patterns by k-means	Number of patterns by SOM
$Q_0$	8215	33715
$Q_1$	1643	6743

Network structure used in the experiments: Number of input neurons equal to the number of attributes of the records in the database (plus the bias input).



**Fig. 2.** Labelling of FV by SOM. (a) Original ROIs. (b) The obtained results of the 3rd partition. (c) The obtained results of the 4th partition. (d) The obtained results of the 6th partition.

Table 3 shows the metaplasticity parameters A and B and the best network structures. Two different criterions to decide for the better network structure and metaplasticity parameters are considered:

1. Metaplasticity parameters: fixing a number of neurons in the hidden layer sufficiently high to presume that the ANN has sufficient processing units to perform the classification, begin to vary the metaplasticity parameters starting with A and finally with parameter B, until we achieve the mentioned value (MSE  $\approx$  0.01) in the minimum number of iterations.
2. Number of neurons in hidden layers: We vary the number of neurons in hidden layers until we achieve the Mean Squared Error (MSE) of approximately 0.01 (metaplasticity parameters are not changed) with the minimum number of neurons without degrading final performance.

**Table 3.** The best network structures and metaplasticity parameters.

Data Set FV <sub>s</sub>	Network Structure			Metaplasticity Parameters		Mean Squared Error
	I	H	O	A	B	
k-means	4	1	1	3	0	0
	4	1	1	3	0	0
SOM	4	1	1	3	0.25	0
	4	1	1	3	0	0

In this work different network structures were used; the activation function for all neurons is sigmoidal with scalar output in the range (0,1); and with the same metaplasticity parameters. A confusion matrix is built to determine the probability of the detection MCs vs. probability of false MCs. Table 4 shows the performance of the classifiers presented in this work.

**Table 4.** Confusion matrices and performance of the classifiers.

Classifier AMMLP	Desired Results	Output Results		Sensitivity (%)	Specificity (%)	Total Classification Accuracy (%)
		MCs	Normal Tissue			
k-means						
Structure1 4 : 15 : 1	MCs	483	0	100	99.67	99.72
	Normal Tissue	8	2466			
Structure2 4 : 12 : 1	MCs	476	7	98.55	99.63	99.45
	Normal Tissue	9	2465			
SOM						
Structure3 4 : 15 : 1	MCs	1328	2	99.84	99.95	99.93
	Normal Tissue	3	6758			
Structure4 4 : 10 : 1	MCs	1317	13	99.02	99.94	99.78
	Normal Tissue	4	6757			

## **4 Conclusions**

In this work two clustering algorithms, k-means and Self Organizing Maps, in order to detect MC clusters in digitized mammograms were used. Clustering algorithms help us to get a better comprehension and knowledge of data with the objective of segmenting the image into different areas (background and MC). After a learning process, the partitional clustering algorithms provide a set of centroids as the most representative elements of each group. As such, clustering algorithms partition the input images in homogeneous areas, each of which is considered homogeneous with respect to a property of interest.

Before applying the clustering algorithms, we applied a digital image processing technique as the image enhancement using mathematical morphology operations in order to improve the contrast between the MC clusters and the background in the ROIs. The mathematical morphology operations based on Coordinate Logic Filters given that the main features of this filters make that their implementation is simple and fast hardware implementation, although in this work only simulation was carried out. Our methodology was tested in different ROIs images, with different kinds of tissue and different shapes of the MCs.

The experimental results show that the proposed method can locate Clusters of Microcalcifications in an efficient way, moreover the method promises interesting advances in Medical Industry.

## **References**

1. Pal, N., Bhowmick, B., Patel, S., Pal, S., Das, J.: A multi-stage neural network aided system for detection of microcalcifications in digitized mammograms. *Neurocomputing*, 71 (13-15), 2625–2634 (2008)
2. Wei, L., Yang, Y., Nishikawa, R.: Microcalcification classification assisted by content-based image retrieval for breast cancer diagnosis. *Pattern Recognition*, 42(6), 1126–1132 (2009)
3. Papadopoulos, A., Fotiadis, D., Costaridou, L.: Improvement of microcalcification cluster detection in mammography utilizing image enhancement techniques. *Computers in Biology and Medicine*, 38(10), 1045–1055 (2008)
4. Bhattacharya, M., Das, A.: Fuzzy logic-based segmentation of microcalcification in breast using digital mammograms considering multiresolution. *International Machine Vision and Image Processing Conference*, 98–105 (2007)
5. Sung-Nien, Y., Kuan-Yuei, L., Yu-Kun, H.: Detection of microcalcifications in digital mammograms using wavelet filter and markov random field model. *Computerized Medical Imaging and Graphics*, 30(3), 163–173 (2006)
6. Guzman-Cabrera, R., Guzman-Sepulveda, J., Torres-Cisneros, M., May-Arrijoja, D., Ruiz-Pinales, J., Ibarra-Manzano, O., Avina-Cervantes, G., Gonzalez-Parada, A.: Digital image processing technique for breast cancer detection. *International Journal of Thermophys*, 34(8), 1519–1531 (2013)
7. Guzman-Cabrera, G., Guzman-Sepulveda, J. R., Torres-Cisneros, M., May-Arrijoja, D., Ruiz-Pinales, J., Ibarra-Manzano, O., Avina-Cervantes G.: Pattern recognition in photoacoustic dataset. *International Journal of Thermophysics*, 34(8), 1638–1645 (2013)

*J. Quintanilla-Domínguez, J. M. Barrón-Adame, J.A. Gordillo-Sosa, J. M. Lozano-Garcia, et al.*

8. Guzman-Cabrera, R., Guzman-Sepulveda, J., Gonzalez-Parada, A., Rosales-Garcia, J., Torres-Cisneros, M., Baleanu D.: Digital processing of thermographic images for medical applications. *Revista de Chimie*, 67(1), 53–56 (2016)
9. Suckling, J., Parker, J., Dance, D.: The mammographic image analysis society digital mammogram database. In: *Excerpta Medica International Congress Series*, 1069. 375–378 (1994)
10. Marcano-Cedeño, A., Quintanilla-Domínguez, J., Andina, D.: Wood defects classification using artificial metaplasticity neural network. In: *35th Annual Conference of the IEEE Industrial Electronics Society, IECON*, 3422–3427 (2009)
11. Andina, D., Álvarez-Vellisco, A., Aleksandar, J., Fombellida, J.: Artificial metaplasticity can improve artificial neural network learning. In: *Intelligent Automation and Soft Computing*, 15(4), 681–694 (2009)