World Automation Congress © 2010 TSI Press.

MICROCALCIFICATION DETECTION APPLYING ARTIFICIAL NEURAL NETWORKS AND MATHEMATICAL MORPHOLOGY IN DIGITAL MAMMOGRAMS

J. QUINTANILLA-DOMINGUEZ[1,2], M.G. CORTINA-JANUCHS[1,2], B. OJEDA-MAGAÑA[1], ALEKSANDAR JEVTIĆ[1], A. VEGA-CORONA[2] AND D. ANDINA[1]

[1]*Group for Automation in Signal and Communications, Technical University of Madrid, Spain*
[2]*Laboratorio de Inteligencia Computacional, LABINCO, University of Guanajuato, Mexico*

ABSTRACT— Breast cancer is one of the leading causes to women mortality in the world and early detection is an important means to reduce the mortality rate. The presence of microcalcifications clusters has been considered as a very important indicator of malignant types of breast cancer and its detection is important to prevent and treat the disease. This paper presents an alternative and effective approach in order to detect microcalcifications clusters in digitized mammograms based on the synergy of the image processing, pattern recognition and artificial intelligence. The mathematical morphology is an image processing technique used for the purpose of image enhancement. A k-means algorithm is used to cluster the data based on the features vectors and finally an artificial neural network-based classifier is applied and the classification performance is evaluated by a ROC curve. Experimental results indicate that the percentage of correct classification was 99.72%, obtaining 100% true positive (sensitivity) and 99.67% false positive (specificity), with the best classifier proposed. In case of the best classifier, we obtained a performance evaluation of classification of Az = 0.9875.

Key Words: Microcalcifications Clusters, Mathematical Morphology, Artificial Neural Networks, Pattern Recognition.

1. INTRODUCTION

Breast cancer is one of the most dangerous types of cancer among women around the world. Early detection of breast cancer is essential in reducing life loss. Currently the most effective method for early detection and screening of breast cancers is mammography. However, achieving this early cancer detection is not an easy task. Although the most accurate detection method in medical environment is biopsy, it is an aggressive, invasive procedure that involves some risks, patient's discomfort and high cost. Microcalcifications (MCs) and masses are two important early signs of the diseases. 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. MCs clusters are primary indicators of malignant types of breast cancer, the detection is important to prevent and treat the disease. But it is still a hard work to detect all the MCs due to the fact that, in mammograms there is a poor contrast between MCs and the tissue around them. Many studies have been focused on the general issue of detection of MCs in mammograms, using several methodologies such as Artificial Neural Networks (ANN), Analysis Wavelet, Support Vector Machines (SVM), Mathematical Morphology, Bayesian Image Analysis Models, High Order Statistic, Fuzzy Logic Systems, etc. Zhao *et al.* [1], analyzed mammograms based on mathematical morphology, in order to obtain suspicious calcification areas. Wirth *et al.* [2], proposed an enhancement algorithm based on morphological analysis. They isolate the breast region and use morphological pre-processing to suppress the background artifacts, then morphological enhancement is used to improve the contrast of the microcalcifications. One of the benefits of morphological contrast enhancement is that it allows fine details to be preserved. Fu *et al.* [3] proposed the segmentation of suspected MC using the top-hat transform, after that they extract images features (textural, spatial and spectral domain). They proved two types of classifiers a General Regression Neural Network (GRNN) and a support vector machine (SVM). Peng *et al.* [4] presented an approach genetic algorithm (GA), in their work extracted features such as shape, texture. Leod *et al.* [5] used a combination of Self Organizing Map (SOM) based clustering with Modified Gramchmidt (MGS) method, using six features, density, mass shape, mass margin abnormality assessment rank an patient age. Veni *et al.* [6] detected MCs clusters using SUSAN edge detector followed by the shape filters, their used features such as perimeter, foreground-background ratio and difference-sum ratio, these features were classified using a feed-forward Neural Network. Tui *et al.* [7] detected MCs using images features such as mean, variance, DCT coefficients and entropy. They used a self-organizing map neural network with fuzzy criterion classifier to classify the

regions with similar characteristics. Zhang *et al.* [9] proposed a bagging-based twin support vector machine (B-TWSVM) to detect MCs, using 164 features. Verma [8], evaluated the impact of multiple clusters in regions of interest (ROIs), using a neural network with six features as inputs. Vega-Corona *et al.* [10]. propose and test a method for the detection of MCs in digital mammography. The method combines selections of Region of Interest (ROI) where MCs were diagnosed, enhanced of the image by histogram adaptive techniques, processing by multiscale wavelet and gray level statistical techniques, clustering and labelling of suboptimal feature vectors applying an unsupervised statistical method base on improved k-means algorithm and a neural feature selector based in a GRNN and detector based in a MLP to finally classify the MCs. In this paper we present a methodology based on Image Processing, Pattern Recognition and Artificial Intelligence for identification of MCs clusters in digitized mammogram images. The method consist in four modules: ROI image selection, ROI image enhancement and feature extraction, data clustering and labeling, finally classification module. Fig. 1 shows the block diagram of our proposed method.

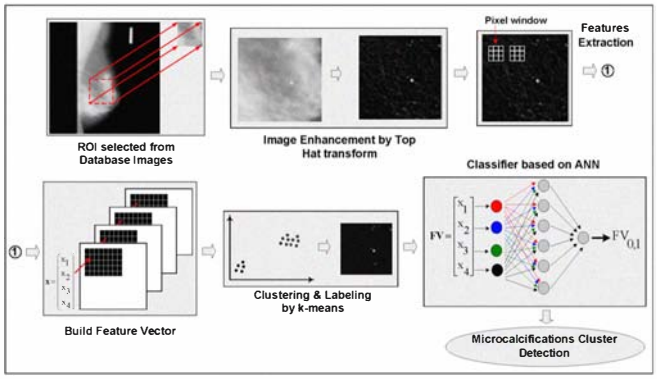


Fig. 1 An overview of the proposed method.

2. ROI SELECTION

The mammograms used to test our method were extracted from a mini Mammographic database provided by Mammographic Image Analysis Society (MIAS) [11]. Each mammogram from the database is a 1024×1024 pixels and with a spatial resolution of $200\mu m$/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 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. IMAGE ENHANCEMENT AND FEATURE EXTRACTION

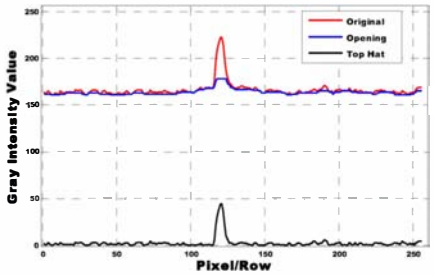


Fig. 2 Application of the Top Hat Transform.

3.1 Image Enhancement

In the field of image processing, image enhancement includes intensity and contrast manipulation, noise reduction, background removal, edges sharpening, filtering, etc. The main goal of mammogram enhancement is to sharpen the edges of ROI, or to increase the contrast between ROI and background. In this work we applied the Top-Hat transform in order to enhancement the ROI image. 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:

$$I_T = I_{in} - [(I_{in} \Theta se) \oplus se] \qquad (1)$$

where, I_{in} input image, I_T transformed image, se structuring element, Θ morphological erosion operation, $\oplus$ morphological dilation operation and $-$ image subtraction operation. $[(I_{in} \Theta se) \oplus se]$ is also known as the morphological opening operation. Fig 2 shows the Top-Hat transform of a row of a ROI image.

3.2 Spatial domain features

MCs clusters 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. These features

are mean and standard deviation, extracted from the 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 clusters of the pixels that correspond to background. The mean and standard deviation are defined by the following equations:

$$I_{\mu}(i,j) = \frac{1}{nxm} \sum_{i=1}^{n} \sum_{j=1}^{m} f(i,j) \quad (2)$$

$$I_{\sigma}(i,j) = \left(\frac{1}{nxm} \sum_{i=1}^{n} \sum_{j=1}^{m} (f(i,j) - I_{\mu}(i,j))^2 \right)^{1/2} \quad (3)$$

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. DETECTION OF MICROCALCIFICATION CLUSTERS

In this work uses the general structure of pattern recognition using a classifier based on ANN which stacks and spatially registers a group of feature images. The MCs are identified applying a clustering method such as, the k-means algorithm. Next, the MCs are represented by an image segmented. The detection and posterior consideration of every MCs in the images can obtain improvements in the results in the process of classification.

4.1 Feature vector creation

We build a Features Vector ***(FV)***, $\mathbf{FV}_s = \{\mathbf{x}^{(qs)}: q_s = 1, \dots, Q_s\}$, where $x^{(qs)} \in \Re^D$ is a *D-dimensional vector* and Q_s is the number of pixels into the image. The ***FV*** set by pixel in ***FVs***, are then clustered and labelled using k-means algorithm. Then, we build a vector $x^{(qs)}$, with four features by pixel $\mathbf{x}^{(qs)} = \{[x_1^{(qs)}, x_2^{(qs)}, x_3^{(qs)}, x_4^{(qs)}]\}$, where $q_s = 1 \dots, Q_s$ and $Q_s = M \times N$,, where $M \times N$ is image size. The pattern vector **x** is represented by a set of four features described as follows: x_1 gray level value of original image, x_2 gray level value of Top Hat image processed x_3 and x_4 correspond to mean and deviation standard of image processed by Top Hat transform.

4.2 Clustering and labelling

FVs are grouped in k-clusters where only one group corresponds to MC (class 1) and other group correspond to some features such as background, vases and normal tissue (class 0). The main idea is that ***FVs*** is clustered en two class and build two sets one for each class $\boldsymbol{FV_1}$ and $\boldsymbol{FV_0}$ around of the prototypes of the class centers $\mathbf{z}^{(1)}$ and $\mathbf{z}^{(0)}$. For this, we applied a well-established clustering method k-means algorithm and then classify the features into ***FVs*** in two class $\boldsymbol{FV_0}$ and $\boldsymbol{FV_1}$, and defined $\boldsymbol{FV_{0/1}}$ respectively in the equation 4.

$$FV_s = \{x^{(q_{0/1})}: q_{0/1} = 1, \dots, Q_{0/1}, x^{(q_{0/1})} \in \mathbb{R}^D\} \quad (4)$$

In this work the used criterions to determine which cluster represent MC group were: minimum number of data clustered into class with maximum gray level value as well as the clusters separability approach by Fisher Linear Discriminant Analysis [12, 13]. We used a between class matrix dispersion (S_b) and intra class matrix dispersion (S_w), these matrix are used to obtain a separability metric between clusters according to equation 5 proposed in [14]:

$$J_d = tr(S_w^{-1} S_b) \quad (5)$$

where, and. $S_b = \sum_{k=1}^{k} n_k (\mathbf{z}^{(k)} - \mathbf{z})(\mathbf{z}^{(k)} - \mathbf{z})^T$, $\mathbf{z} = 1/M \sum_{q=1}^{M} \mathbf{x}^{(q)}$ and $M = \sum_{k=1}^{K} n_k$. Where k is the number of clusters, $\mathbf{x}^{(qk)}$ is the vector into cluster Q_k and n_k is the number of vectors into k cluster, M is the total features vector. The results of clustering are represented as a segmented image of binary form.

5. MICROCALCIFICATIONS CLASSIFICATION

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 Normal class (0) or MCs class (1). For this purpose, we applied Feed Forward Neural Network (FFNN). FFNN known as Multi-Layer Perceptron (MLP) is the most popular used in many practical applications. FFNN is a type of supervised learning. Knowledge is acquired

by the network through a learning process known as the Back Propagation (BP) algorithm. The BP algorithm has emerged as the workhorse for the design of a special class of layered FFNN. A FFNN has an input layer of source nodes and an output layer of neurons these two layers connect the network to the outside world. In addition to these two layers, the multilayer perceptron usually has one or more layers of hidden neurons, which are so called because these neurons are not directly accessible. The hidden neurons extract important features contained in the input data. Using supervised learning, these networks can learn the mapping from one data space to other using examples. The terms BP refers to the way that error computed at the output side is propaged backward from the output layer, to hidden layer, and finally to the output layer, details of this method can be found in [15]. For the classifier design, one has to obtain a number of nodes in each layer, number of hidden layers and the type of the training function. The number of hidden layers is usually chosen based on training of the network under various configurations. In many cases, a single hidden layer is sufficient, but the addition of the second hidden layer improves network's prediction capability due to nonlinear separability property of the network.

6. EXPERIMENTAL RESULTS

In this section, the experiments results obtained of each stage of the methodology are presented. In first stage, we selected several ROI images from mammograms with dense tissue and the presence of MCs. Next, the morphological Top-Hat transform is used in order to enhancement the ROI image, with the goal of to segment objects that differ in brightness from the surrounding background. In this work, we apply the same structuring element at different sizes, but the best results in this stage were obtained with the 9×9 size. Fig. 3 shows a original ROI image processed by Top Hat transform, as well as topographical relief of each image. 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. Five size windows were applied. The correlation analysis was made in order to find the best pixel block window, finally according with the obtained results, we have chosen a 5×5 pixel block window. In the next stage, we

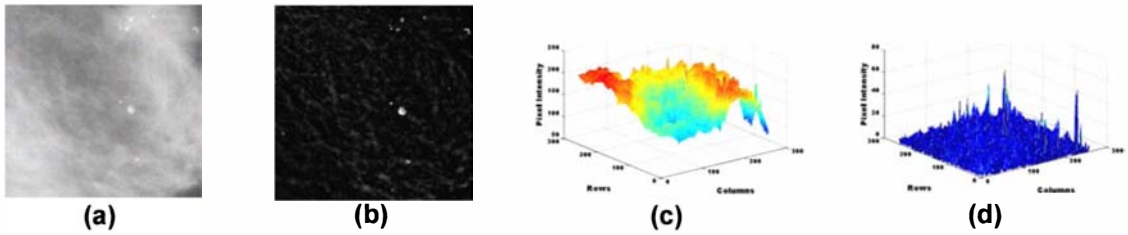
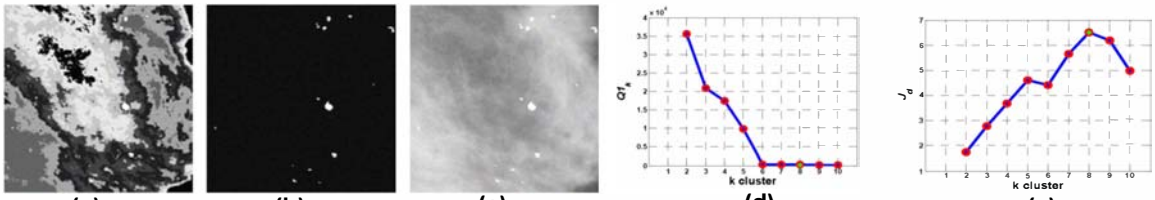


Fig. 3 (a) Original ROI. (b) Enhancement ROI. (c) Topographic relief of original ROI. (d) Topographic relief of enhancement ROI.

build a $\boldsymbol{FV}$ with four features by pixel, where each feature corresponds to gray level intensity of the original ROI image, the gray level to the enhancement ROI image and the mean and standard deviation of the enhancement image. After a clustering method was implemented in order to obtain the label of the patterns of the $\boldsymbol{FV}$, $\boldsymbol{FV_s}$ is clustered in k-clusters according to similarity of pixels features in the images. The k-means algorithm was used to clustered the $\boldsymbol{FV}$. Where the basic idea is obtain one cluster that corresponds to MCs ($\boldsymbol{FV_1}$) and other ones correspond to vases, background and tissue, generally the cluster that corresponds to normal tissue ($\boldsymbol{FV_0}$). This we can represent as a segmented image. In figure 4(a) shows a ROI image segmented to 8 classes, although it could continue segmenting, we can see that there is a convergence of class center thus obtaining a new label of the $\boldsymbol{FV_s}$. Fig. 4(b), shows the ROI image segmented into 2 classes where one class represent a MCs and the other one normal tissue, Fig. 4(c) shows the MCs found by our methodology and they are displayed superimposed on the original ROI. Fig 4(d) shows a number of labeled patters as MC *vs.* number of clusters *k*. Another criterion for determining the number of clusters is to calculate the maximum separability between classes by Fisher Linear Discriminant Analysis, Fig 4(d) shows the maximum separability evolution *vs.* number of clusters *k*. From the obtained results in clustering stage, we obtained the labeled sets $\boldsymbol{FV_0}$ with 588181 patterns and $\boldsymbol{FV_1}$ 1643 patterns. Due to the large amount of patterns of the class that not belong to MCs respect to the number of patterns that belong to the class of MCs a balancing was performed, then we obtain two subsets with 1643 patterns to the class $\boldsymbol{FV_1}$ and 8215 patterns to $\boldsymbol{FV_0}$. In this work we used patterns extracted of the set $\boldsymbol{FV_s}$ in order

to train and test our the classifier, for this case we used the 70% of the data to train, 6901 samples, of which 1160 are MCs and 5741 are normal tissue. 30% of the data were used to test, 2657 samples, of which 483 are MCs and 2474 are normal tissue. In this work we apply different network structures, of which the best obtained results were with the following structure 4:9:1 fully connected with a sigmoidal function in each hidden node.



(a) (b) (c) (d) (e)

Fig 4. (a) Segmented ROI to 8 classes. (b)Binary ROI image. (c)Detected MCs in original ROI. (d)Number of labeled patterns as MC vs. number of clusters. (e)Maximum separability evolution vs. number of clusters.

The node output is thresholded to obtain a output between [0,1]. With the following structure 4:12:1 we obtained different results with the previous structure. Next, we build a confusion matrix to determine the probability of the detection MCs (true positive rate, TP) *vs* probability of false MCs (false positive rate, FP). In Table I shows the performance of the classifiers presented in this work. The performance of the proposed method is evaluated by means of the ROC (Receiver Operating-Characteristics) curve analysis. The ROC curve is a plot of the *Sensitivity vs*. *Specificity* for the different possible cut-points of a diagnostic test. Fig 5 shows the ROC curves and the area under the curve (A_z) to the classifiers with the structure 1 and structure 2, respectively.

Table I. Confusion matrices and performance of the classifiers presented.

Classifier	Desired Results	Output Results		Sensitivity (%)	Specificity (%)	Total Classification Accuracy (%)
		MCs	Normal Tissue			
Structure 1 4:9:1	MCs	483	0	100	99.67	99.72
	Normal Tissue	8	2466			
Structure 2 4:12:1	MCs	476	7	98.55	99.63	99.45
	Normal Tissue	9	2465			

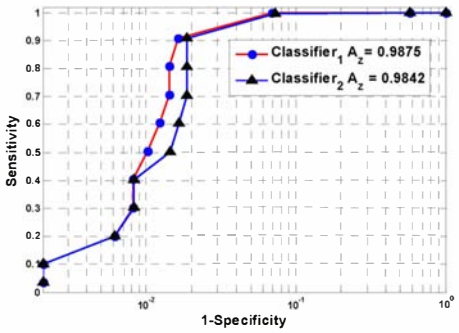


Fig 5. ROC curves of the classifiers.

Table II. Comparison of this work with other methods presented.

Autor	Method	A_z
A. Vega-Corona *et al.*(2003)	GRNN	0.954
J.C. Fu *et al.*(2005)	GRNN/SVM	0.978 /0.98
Y. Peng *et al.*(2006)	KD-GA	0.952
X. Zhang *et al.* (2009)	B-TWSVM	0.9627
This work	**FFNN(BP)**	**0.9875**

Table II shows the A_z values obtained in previous works, in order to compare the proposed methodology in this work.

7. CONCLUSIONS

This paper proposed an alternative method for Microcalcifications clusters detection in digitized mammograms using image processing, data clustering and ANN. Feature extraction in ROIs images is an important factor for the MCs detection, due to the low contrast in these images type. We successfully applied the Top Hat transform in order to enhance the ROI image; additionally, the mean and the standard deviation values were used as complementary information about the areas of interest.

Clustering algorithms help us to get a better comprehension and knowledge of the analysed data with the objective to segment the image into different areas according to the problem at hand. In this work, we propose an ANN-based classifier. The ANNs have been used with success in different research areas. The classifier plays an important role in our methodology, because the ANN can learn from the data patterns contained in the training set and conduct complex decision-making. Experimental results presented in this work indicate that our method is an interesting alternative for MCs detection in mammograms in comparison with the methods proposed by the other researchers.

ACKNOWLEDGEMENTS

The authors wishes to thank to The National Council for Science and Technology (CONACyT), The Secretariat of Public Education (SEP) and Government of Mexico and Group of Automation in Signal and Communications (GASC) of Technical University of Madrid for their contribution.

REFERENCES

1. D. Zhao, M. Shridhar and D.G. Daut, "Morphology on detection of calcifications in mammograms", *IEEE International Conference on Acoustics, Speach and Signal Processing, ICASSP-92*, San Francisco, USA, 1992, pp. 129-132, 1992.
2. M. Wirth, M. Fraschini and J. Lyon, "Contrast enhancement of microcalcifications in mammograms using morphological enhancement and non-flat structuring elements", *17th IEEE Symposium on Computer-Based Medical Systems, CBMS 2004.* Bethesda, MD, USA, 2004 , pp.134-139.
3. J.C. Fu, S.K. Lee, S.T.C. Wong, J.Y. Yeh, A.H. Wang and H.K. Wu, "Image segmentation feature selection and pattern classification for mammographic microcalcifications", *Computerized Medical Imaging and Graphics(29)*, Elsevier, 2005, pp.419-429.
4. Y. Peng, B. Yao and J. Jiang, "Knowledge-discovery incorporated evolutionary search for microcalcification detection in breast cancer diagnosis", *Artificial Intelligence in Medicine 37(1)*, Elsevier, 2006, pp.43-53.
5. P.Mc. Leod, B. Verma and R. Panchal, "Combining SOM based Clustering and MGS for Classification of Suspicious Areas within Digital Mammograms". *3rd International Conference on Intelligent Sensors, Sensor Networks and Information, ISSNIP '07*, Melbourne, Australia, 2007,pp. 413-418.
6. G. Veni, E.E. Regentova and A.K. Mandava, "A New Method of Detecting Microcalcification Clusters for Computer Aided Digital Mammography", *19th International Conference on Systems Engineering, ICSENG '08,* Las Vegas, USA, 2008, pp.532-537.
7. C.M. Tiu, T.L. Jong and C.W Hsieh, "Self organizing map neural network with fuzzy screening for micro-calcifications detection on mammograms", *IEEE Conference on Soft Computing in Industrial Applications, SMCIa '08*, Muroran, Japan, 2008, pp.421-425.
8. B. Verma, "Impact of multiple clusters on neural classification of ROIs in digital mammograms". *International Joint Conference on Neural Networks, IJCNN '09*, Atlanta, USA, 2009,pp. 3220-3223.
9. X. Zhang and H. Xie, "A New Approach for Clustered Microcalcifications Detection". *Asia-Pacific Conference on Information Processing, APACIP '09*, Shenzhen, China, 2009, pp.322-325.
10. A.Vega-Corona, A.Álvarez, and D.Andina, "Feature Vectors Generation for Detection of Microcalcifications in Digitized Mammography Using Neural Networks", *Artificial Neural Nets Problem Solving Methods,* LNCS Springer Berlin-Heidelberg, Vol.2687, 2003, pp. 583-590.
11. J.Suckling, J.Parker and D. Dance, "The mammographic image analysis society digital mammogram database. Exerpta Medica". *International Congress Series,* Vol.1069, 1994, pp. 375-378.
12. R.O. Duda, P.E. Hart and D.G Stork. "Pattern Classification", Wiley-Interscience, 2nd ed. 2001.
13. A.Vega-Corona, A.Sanchez-Garcia, M. Gonzalez-Romo, J. Quintanilla-Dominguez, J.M Barron-Adame, "Contextual and Non-Contextual Features Extraction and a Selection Method for Microcalcifications Detection". *World Automation Congress, WAC '06,* Sevilla, Spain, 2006, pp.1-6.
14. K. Fukunaga. "Introduction to Statistical Pattern Recognition", Academic Press, 2nd ed., 1990.
15. I.A.Basheer and M.Hajmeer, "Artificial neural networks: fundamentals, computing, design, and application", *Journal of Microbiological Methods 43(1),* Science Direct, 2000, pp. 3-31.