Automated Detection and Classification of Clusters of Microcalcifications in Mammograms

Alain Tiedeu* *GRETMAT, LETS, ENSP, Univ. Yaoundé I. BP 8390, Yaoundé Cameroon* alain_tiedeu@yahoo.fr

Aude D. Kentsop *GRETMAT, LETS, ENSP, Univ. Yaoundé I. BP 8390, Yaoundé Cameroon* audekentsop@yahoo.fr

Christian Daul *UMR 7039 CRAN-CNRS-UHP-INPL, 2 av. Forêt de Haye, Vandoeuvre France* cdaul@ensem.inpl-nancy.fr

Pierre Graebling *UMR 7005, LSIIT, Bd S.Brant, 67400, Illkirch-Graffenstanden. France* pierre.graebling@ensps.u-strasbg.fr

Martin Kom *GRETMAT, LETS, ENSP, Univ. Yaoundé I. BP 8390, Yaoundé, Cameroon* kommart@yahoo.com

* : corresponding author

Abstract

Computer-aided-diagnosis (CAD) systems can be very useful as « second opinion » in case of mammography screening programs where large volumes of mammograms are produced, of which only a few have abnormalities to be spotted out.

We propose a full CAD tool for mammography. It comprises two modules. The first module detects microcalcification clusters (MCCs) present on the mammograms. It is based on a two-step algorithm. The first step segments the images. In the second step, candidates which are not true MCs are discarded. Only MCs belonging to a cluster are retained. The second module classifies the detected MCCs into benign and malignant types. Testing both modules, the first detected 85.65% of individual MCs and 100% of the MCCs present on the mammograms and the second yielded a sensitivity of 100%, a specificity of 75% and good classification rate of 85.29% .

Keywords: computer-aided-diagnosis, detection, microcalcification cluster, neural network, classification.

1. Introduction

Mammography is nowadays a widespread technique for early detection of breast cancer . Microcalcifications clusters (MCCs) are generally an early sign of breast cancer. Microcalcifications (MCs) are tiny granular deposits of calcium which can be seen on a mammogram. Detecting and classifying impalpable malignant MCs within the breast can improve survival rate of breast cancer patients. Computer-aided diagnostic systems which would assist radiologists detecting and classifying such MCs are therefore very useful. A number of works in the literature have been devoted to developing such systems. An excellent review of the subject was provided by Cheng et al.[1]. Each approach has its own strengths and weaknesses. As in our paper two parts generally emerge: the first deals with detection and the second with classification of MCs.

The detection of the MCs' part of the system has been attempted by many authors. For example, methods based on wavelet and other multi-scale approaches [2], fuzzy logic [3], neural networks [4], texture analysis [5], statistical approaches [6] or genetic algorithm [7] have been suggested. All these methods can be roughly divided into two approaches. In the first approach, there are often a pre-processing (signal to noise enhancement) and a segmentation step. Both are designed so that as many MCs as possible are detected without taking into account signals which lead to false positives (FPs). Some authors of whom Strickland et al. [8] are examples, used this approach. In the second approach, the eventual pre-processing step and the MC segmentation are developed to detect as many MCs as possible while minimizing FPs. Pre-processing and detection are carried out either sequentially or simultaneously. Global or local tresholding techniques are then used to binarized the image. Chan et al. [9] are some of the authors who used this approach.

In order to detect the MCs, we developed an algorithm that combined the advantages of the two

approaches described above. The segmentation step which is designed to find as many MCs as possible while minimizing FPs (approach 2) is followed by a selection step (approach 1) eliminating unavoidable FPs from the segmentation. MCCs being more relevant for diagnosis, we made sure that only MCs belonging to clusters were kept. Sub-section II.1.1. gives more details on that.

The classification system generally consists of 3 steps in the literature: computation of appropriate features, selection of the most meaningful set of features for the classification, feeding the features into the classifier for malignant analysis of the lesions. The literature offers two main methods of features' selection: the genetic algorithm (GA) and step wise or iterative method (IM). The classifiers play a major role in the CAD system and four kinds of classifiers are mostly used: artificial neural networks (ANN) ([10]), k-nearest neighbour classifiers ([11]), Bayesian belief networks ([12]), and binary decision tree [13]. An analysis of results provided by different authors proves that textural features seem to be the best choice for feature sets [14] on one hand. On the other hand ANN, on the average, gave better results than the three other classifiers [1]. Finally, due to complexity of implementation and time consumption of GAs, the IM is generally preferable.

We therefore proposed a classification module using a set of textural features derived from the MCCs. Using the IM, the best set of feature from the thirteen previously computed was selected. And finally, using a back propagation ANN made of an input layer, one hidden layer and an output layer, the detected MCCs were classified.

The main advantages and differences of the proposed system as compared to those present in the literature are on one hand the fact that we developed a novel detection algorithm efficiently combining the two approaches of MCs detection techniques . On the other hand, we used an adapted IM for selection of features and some textural features hitherto unused for MCCs classification. In the following section, details of the algorithms used in each module are given.

2. Details of the algorithms

2.1. Detection module

2.1.1. Image segmentation. The principle of the detection algorithm is given in figure 1. The principle of our segmentation algorithm is identical to that of Chan et al [9], in the sense that the original image (*OI(x,y)*) is smoothed (*SI(x,y)*) and subtracted from an image in which the contrast has been enhanced (*EI(x,y)*). In Chan's algorithm, linear filters called "matched filters" were used to enhance the contrast between the calcifications and the mammary tissues while either median filters or "contrast reversal filters" allowed them to smooth the original images. Contrast enhancement was obtained in our case through an algorithm proposed by Gonzales and Wintz [15] which is described below.

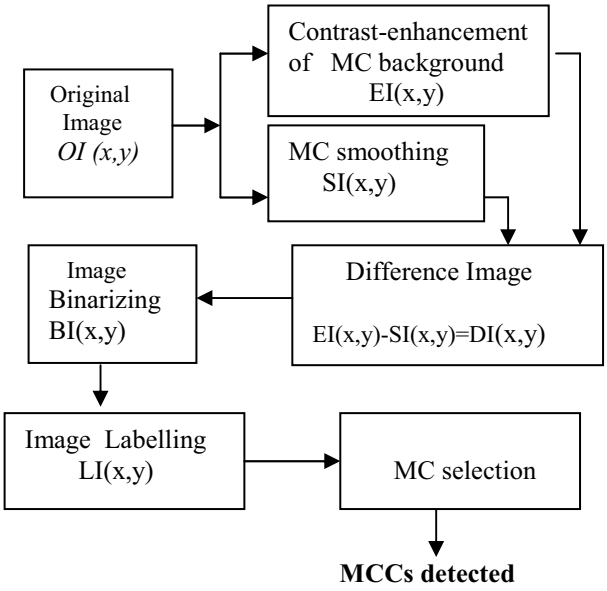

Figure 1. Principle of the detection algorithm

Let us consider a rectangular neighbourhood (of width *w* and height *h*) centred on a pixel of coordinates *(i,j)* in *OI(i,j)*. The mean grey value and the standard deviation in this neighbourhood are $\overline{m}_{i,j}$ and $\sigma_{i,j}$ respectively. The value of the pixel of coordinates *(i,j)* of the contrast enhanced-image is :

$$EI(i,j) = k_{scale} \frac{\overline{m}_{OI}}{\sigma_{i,j}} (OI(i,j) - \overline{m}_{i,j}) + \overline{m}_{i,j} \quad (1)$$

$\overline{m}_{OI}$ and k_{scale} are respectively the mean of grey-level value computed with all the pixels of *OI(x,y)* and a scaling factor introduced to keep the resulting grey-values in a given range.

The term $(OI(i,j) - \overline{m}_{i,j})$ is large if *OI(i,j)* belongs to a MC and is on the average smaller if *OI(i,j)* corresponds to mammary tissue. Moreover, if the texture of the mammary tissue is very pronounced, the term $\frac{\overline{m}_{OI}}{\sigma_{i,j}}$ is small ($\sigma_{i,j}$ being large). Since $\frac{\overline{m}_{OI}}{\sigma_{i,j}}$ modulates $(OI(i,j) - \overline{m}_{i,j})$, the latter difference decreases if all considered pixels are on mammary tissues. This algorithm therefore amplifies the strong

grey-level slopes which generally correspond to MCs while enhancement of mammary tissue is limited.

We used gaussian filters to smooth the original image. By subtracting pixelwise the smoothed-image from the contrast-enhanced image, the background is strongly attenuated and the MCs more accentuated.

$$DI(i,j) = EI(i,j) - SI(i,j) \quad (2)$$

The difference-image *DI(i,j)* obtained here above is binarized using a local adaptive tresholding algorithm [9]. Its principle is the following. For a rectangular window centred on a pixel of coordinates *(i,j)* and of width w_{bin} and height h_{bin} , the mean grey-value $\overline{m}_{i,j}^{bin}$ and the standard deviation $\sigma_{i,j}^{bin}$ are computed. The pixel *(i,j)* in the resulting image is *(BI(i,j))* .

$$BI(i,j) = 1 \quad \text{if} \quad BI^{pix}(i,j) > \overline{m}_{i,j}^{bin} + k\sigma_{i,j}^{bin}$$

$$BI(i,j) = 0 \quad \text{otherwise} \quad (3)$$

k is a pre-selected integer and $\sigma_{i,j}^{bin}$ is the standard deviation computed using the grey-value of the pixel coordinates and $\overline{m}_{i,j}^{bin}$ the mean grey-level value in the window. In this paper, w_{bin} and h_{bin} were chosen equal to 71 since this size of the window is large enough to include "a representative mammary texture". The binarized-image was then labelled using a classical eight neighbourhood algorithm and yielded *LI(i,j)*.

2.1.2. Microcalcification and cluster of microcalcification selection. Each region *LI(x,y)* with the same label represents a potential MC. In order to distinguish region which are not MCs (false positives) from those which are truly MCs (true positives), a set of moment-based geometrical features were computed. They have been listed below (a-j).

a. Central moment of order p+q : $cm_{p,q}$

$$cm_{p,q} = \sum_{j=0}^{h}\sum_{i=0}^{w}(x_i - \overline{x})^p (y_j - \overline{y})^q f(x_i, y_i) \quad (4)$$

b. Volume in terms of grey-level

$$v = cm_{0,0} \quad (5)$$

c. Elliptical disc (of same volume and second order moment with the MC) and with orientation θ major axis ga and small axis pa

$$\theta = \frac{1}{2}\tan^{-1}\frac{2cm_{1,1}}{cm_{2,0} - cm_{0,2}}$$

$$(ga, pa) = \sqrt{2\frac{cm_{2,0} + cm_{0,2} \pm \sqrt{(cm_{2,0} - cm_{0,2})^2 + (2cm_{1,1})^2}}{cm_{0,0}}} \quad (6)$$

d. Radius of gire

$$rg = \frac{cm_{2,0} + cm_{0,2}}{cm_{1,1}} \quad (7)$$

e. Grey-level concentricity

$$c = \frac{cm_{2,0} + cm_{0,2} + \sqrt{(cm_{2,0} - cm_{0,2})^2 + (2cm_{1,1})^2}}{cm_{2,0} + cm_{0,2} - \sqrt{(cm_{2,0} - cm_{0,2})^2 + (2cm_{1,1})^2}} \quad (8)$$

f. Area of the MC

a = number of pixels above the threshold

g. Perimeter of the MC

p = number of pixels on the border of the MC

h. Circularity factor

$$cf = \frac{p^2}{4\pi a} \quad (9)$$

i. Grey-level slopes around the MC

S_i : i^{th} slope computed with grey-levels of two neighbouring pixels located on each side of the border of the MC

j. Mean grey-level slope around the MC border

$$\overline{S} = \frac{1}{p}\sum_{i=1}^{p} S_i \quad (10)$$

Typical values for MCs for each of the features were found empirically. From these values, a simple multi-threshold algorithm was used to discard the FPs. A further selection procedure was carried out – based on the fact that a cluster of MCs is defined by a minimal density of 5 MCs/cm^3 – to only retain MCs which belonged to a cluster.

2.2. Classification module

The usefulness of co-occurrence matrices texture measure in differentiating malignant and benign breast tissues was demonstrated by analysis of mammographic MCs [16]. We therefore used textural features extracted from our ROIs (MCCs) as information fed into our classifier. The features used are described in the next sub-section.

2.2.1. Feature computation. The co-occurrence matrix element, *MATCOOC(i,j)* is the joint probability of the occurrence of grey-levels *i* and *j* for pixel pairs which are separated by a distance *d* and at a direction *θ*. From work by Chan et al. [16], it is known that there is no significant dependence of the discriminatory power of the texture features on the direction of the pixel pairs for mammographic textures on one hand. We therefore used the direction $\theta = 0°$. On the other hand, in order to obtain fine details of the texture, we used $d = 4$. The co-occurrence matrix was computed on the region of the MCC after which the following 13 texture features were extracted [15]: Correlation , entropy , energy, Inertia, Inverse difference moment,

Sum average, Sum entropy, Sum variance, Difference entropy, Variance, Inverse difference, Information of correlation 1, Information of correlation 2, These features were chosen on the basis of their discriminatory power from literature analysis [16].

2.2.2. Classifiers. Of the diverse classifiers used in the literature, back-propagation ANN seemed to give the best results [1]. An ANN is a parallel, distributed information processing structure consisting of elements interconnected by directional connections. A neural element carries out local operations. ANN can efficiently learn non-linear mappings through examples contained in a training set, and conduct complex decision making. The back-propagation is probably one of the most well research training algorithms. Its main strength resides in back-propagation of the errors from the output, in order to minimize them. We used a very simple back-propagation ANN with an input layer, one hidden layer and an output layer.

2.2.3. Feature selection. We used an iterative method to select the best set of features for the classification of the MCCs. We started with a pool consisting of the 13 features listed here above and carried out the classification with it. The rate of good classification (GC) was noted. Then the influence of each of the feature was evaluated by removing each one of them at its own time and carrying out the classification with the rest. The best set of 12 features was retained. On this set, the same procedure of evaluating the influence of each feature was done and the best set of 11 features retained. We continued this process down to the best set of 3 features after which the classifier gave erroneous results. The best set out of the 10 sets retained was the one which gave the best rate of GC. The rate GC must be understood as the percentage of diagnostic decision (given by the CAD system) that proved to be correct when compared with the biopsy-proven or histological-proven results.

3. Results

3.1. Database

In order to test our two modules, we used a database of 66 images. These contained a total of 59 MCCs and 683 MCs. The images, of size 512 x 512 pixels, encoded on 16 bits and of resolution 98µm/pixel were obtained from the cancerology centre "Centre Alexis Vautrin (CAV)", Nancy, France.

3.2. Results of the detection module

Of the sixty-six images of the data set, eighteen were chosen to study the efficiency of the detection algorithm according to parameters of the contrast enhancement algorithm (window of width w_{enth} and height h_{enth}) and of the smoothing filter (standard deviation σ_{smo} of the Gaussian function).

The trade-off between true positive (TP) cluster rate and false positive (FP) clusters per image is usually represented by free-response receiver operating characteristics (FROC). Such FROC-curves were plotted for several sets of w_{enth}, h_{enth} and σ_{smo} parameters. These curves were built by varying the number of standard deviation (k = *5, 4, 3, 2.5* and *2.1*) in formula (3) (figure 2).

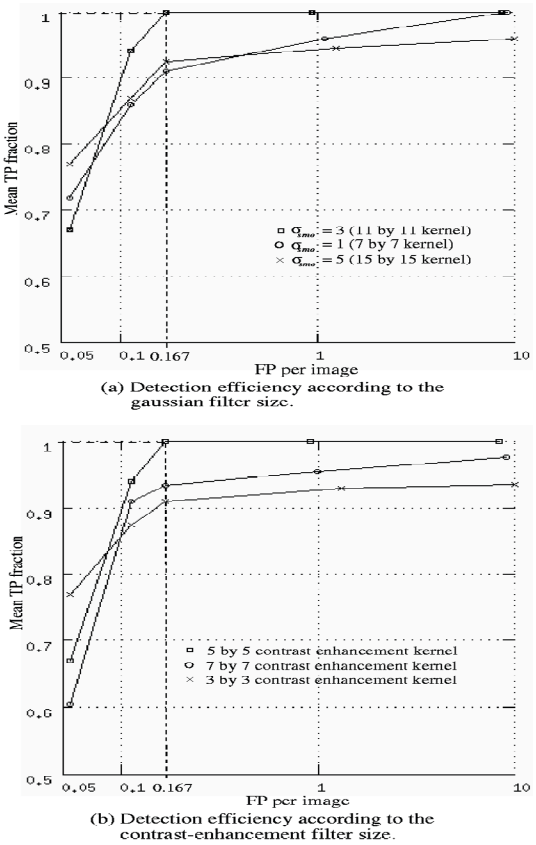

Figure 2. FROC-curves for the detection module of our CAD system

As can be seen, the best results were obtained with the triplet (w_{eath}=h_{enth}=*5* and σ_{smo}=*3*) only 0.16FP/image were detected with these parameters. The detection algorithm was also applied to 48 remaining images of the database. All the clusters present were detected and we registered a total of 5FP, which is a rate of 0.11 FP/image. Finally, the number

of MCCs present, the success rate of the detection module was 100% with 0.121 FP/image. Considering individual MCs, the success rate was 85.65% with 2.50 FP/image. The algorithm was therefore validated by the FROC-curves above (Figure 2). For most algorithms of the literature, the evaluation is given in terms of MCC. Some of them ([2], [3], [5], [17]) have a detection rate around 100% like ours but exhibits a FP rate of at least 0,7 FP cluster /image which is larger than our FP rate. In the recent literature, Karssemeijer et al., [6] have evaluated their algorithms in terms of individual MCs. They obtained a FP rate of 2,97MC/image (which is larger than the 2,5FP/image yielded by our algorithm).

3.3. Result of the classification module

Thirty-four images for which histological-proven results were available were used to test this module. Their characteristics were fed into the ANN.

Table 1. Performance of our classifier

Type of image (according to biopsy or histology)	Cancerous		Non-cancerous		Total
Classification results by the CAD system	Cancerous (TP)	Non-cancerous (FN)	Non-cancerous (TN)	Cancerous (FP)	
Number of images as classified by the CAD system	14	0	15	5	34
Percentages	100% (sensitivity)	0%	75% (specificity)	25%	
Good classification rate					85,29%

The best GC rate was obtained for the following set of nine features: correlation, entropy, energy, sum average, sum entropy, difference entropy, variance, inverse difference, information measure of correlation 1. The results are recapitulated in table 1.

A decision for classification result can be one of four possible categories: true positive (TP), true negative (TN), false positive (FP), false negative (FN). FN and FP are two kinds of errors. A FN error implies that a true cancer was classified non-cancerous region (benign) and a FP error occurs when a non-cancerous region has been classified cancerous. A TP decision is a correct judgement of an actual cancer, and a TN decision means a non-cancerous image was correctly labelled. These lead to :

sensitivity = TP/(TP+FN) (11)

specificity = TN/(TN+FP) (12)

From (11) and (12), table 1 therefore exhibits a sensitivity of 100% and a specificity of 75%. The overall GC rate is 85.29%. These results prove that no cancer is missed, which is good for patient survival rate. On the other hand a 75% rate specificity means that a quarter of patients run the risk of undergoing unnecessary biopsies. This may be improved by injecting some *a priori* knowledge during the building of ANN.

Some authors (Chan et al [16]) proposed systems with very high sensitivities like ours. But they are generally followed, like in our case, by low specificities which is a dilemma in the global evaluation of the system. For a fixed discrimination, the sensitivity and specificity of a diagnostic system will depend on the particular confidence threshold that the observer or the computer diagnostic uses to partition continuously distributed perception of evidence into categorical decision.

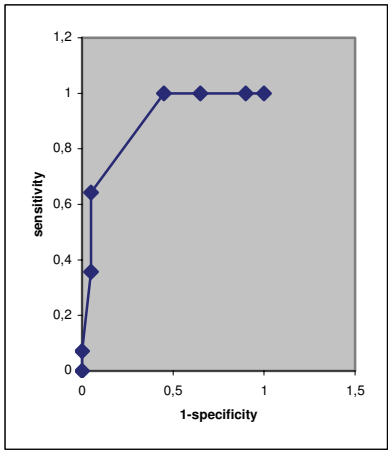

Figure 3. ROC-curve for the classification module of the CAD system

Sensitivity and specificity change when the confidence threshold is changed. A way to solve the problem is to plot Receiver Operating Characteristics (ROC)-curves, which represent the trade offs between sensitivity and specificity and therefore describe the inherent discrimination capacity of a system. A non parametric estimation of the ROC-curve from our results yielded the curve in figure 3.

The area under the ROC curve (*Az*) is an important criterion for evaluating diagnostic performance. The area under the ROC-curve was found equal to 0,89. It is in the order of the highest values in the literature.

4. Conclusion

In this paper a Computer-Aided-Diagnosis system for early detection of breast cancer was presented. MCCs which are often early signs of breast cancer can be detected and classified by this system. The detection module of the system efficiently combined the two main approaches of systems of the same kind in the literature. The high sensitivity (100%) of the system which means that all cancers are detected, is balanced by a relatively low specificity (75%). The latter may probably be improved by injecting some *a priori* knowledge during ANN design. Overall, although the system was tested with a small database, it provided an $Az = 0,89$ which is a performance good enough to suggest the trial of this system at a larger scale.

5. Acknowledgements

The authors wish to thank the French association "La ligue contre le cancer (delegation Meurthe et Moselle)", IRD (Institut de Recherche pour le Développement), AUF (Agence Universitaire de la Francophie), the Abdus Salam International Centre for Theoretical Physics (ICTP) and The Swedish International Development Agency (SIDA) for their financial support. They also thank Dr. B. Boyer, J. Stinès and team of the CAV Cancerology Centre for providing the mammograms and the histological-proven results.

6. References

01. M.D.Cheng, X.Cai, X.Chan, L Hu, X Lou. Computer-aided detection and classification of microcalcifications in mammograms: a survey. Pattern recognition, 2003, Vol. 36, pp 2967-2991.
02. T.Netsch and H-O. Peitgen. Scale-space signatures for the detection of clustered microcalcifications in digital mammograms. IEEE Trans. on Medical Imaging, September 1999, Vol. 18(9), pp 774-786,.
03. H-D. Cheng, Y.M. Lui, and R.I.Freimanis. A novel approach to microcalcification detection using fuzzy logic technique. IEEE Trans. On Medical Imaging. 1998, Vol. 17(3), pp442-450.
04. S-C. Lo, H. Li, J.-S. Lin, A. Hasegawa, O.Tsujii, M.Freedman, and S. Mun. Detection of clustered microcalcifications using fuzzy modelling and convolution neural network. In Murray H. Loew and Kenneth M. Hanson, editors, Proc. SPIE Medical Imaging 1996: Image Processing, vol. 2710, pp8-15.
05. JK Kim, HW Parker. Statistical features for detection of microcalcifications in digitized mammograms. IEEE Trans. on medical imaging, 1999, vol. 18, pp231-238.
06. N. Karssemeijer. Adaptive noise equalization and recognition of microcalcification clusters in mammograms. Int. Journal of Pattern Recognition and Artificial Intelligence, 1993, Vol. 7(6), pp1357-1376.
07. E.Tao, C.Ornes, and J.Sklansky. Automated detection of microcalcifications in digital mammography. In K.M. Hanson, editor, Proc. SPIE medical Imaging 1998: Image Display, vol. 3338, pp1450-1458.
08. R. Strickland and T.Theodosiou. Fuzzy system for detecting microcalcifications in mammograms. In Bosacchi B.,D. Fogel, and J. Bezdek, editors, Proc. SPIE . Applications and Science of Neural Networks, Fuzzy Systems, and Evolutionary Computation, October 1998, vol. 3455, pp317-331.
09. H.P. Chan, K.Doi, S. Galhotra, C. Vyborny, and H.MacMahon. Image feature analysis and computer-aided diagnosis in digital radiography. I. Automated detection of microcalfications of mammography. Medical Physics, 1987, vol. 15(4), pp534-548.
10. A.P. Dhawan, Y. Chitre, M. Moskowitz, Artificial neural network based classification of mammographic microcalcifications using image structure features, Proc. SPIE, 1993,vol.1905,pp820-831.
11. D. Kramer, F. Aghdasi, Classification of v in digitised mammograms using multiscale statistical texture analysis, Proceeding of the South African Symposium on Communications and Signal Processing, September 7-8, 1998, pp.121-126.
12. F.V.Jensen, An Introduction to Bayesian Network, Springer, New York, NY, 1996.
13. S. Bothorel, B. B. Meunier, S.Muller, A fuzzy logic based approach for semiological analysis of microcalcifications in mammographic images, Int. J. Intelligent Systems,1997, vol. 12, pp819-848.
14. A.B. Tiedeu. Contribution aux outils de diagnostic assiste par ordinateur du cancer du sein en 2D et 3D. Thèse de doctorat d'état en Physique, Université de Yaoundé I, 2004.
15. R.C. Gonzalez and P. Wintz. Digital Image Processing (2nd Edition). Addison-Wesley Publishing Company, 1987.
16. H.P. Chan, B. Sahiner, K.L. Lam, N.Petrick, M.A Helvie, M.M. Goodsitt, DD Adler. Computerized analysis of mammographic microcalcifications in morphological and texture feature spaces. Medical physics, 1998, vol. 25(10): 2007-2019.