CLASSIFICATION OF MICROCALCIFICATIONS IN DIGITAL MAMMOGRAMS USING CASE-BASED REASONING

Joan Martí [a], Josep Español [b], Elisabet Golobardes [c],
Jordi Freixenet [a], Rafael García [a], and Maria Salamó [c]

[a] Computer Vision and Robotics Group
Institute of Informatics and Applications. University of Girona
Avda. Lluís Santaló s/n 17071 Girona, Catalonia (Spain)
e-mail: {joanm,jordif,rafa}@eia.udg.es

[b] Department of Oncology. University Hospital ``Dr. Josep Trueta"
Avda. de França s/n 17007 Girona, Catalonia (Spain)
e-mail: {jesp}@arrakis.es

[c] Computer Science Department
Enginyeria i Arquitectura La Salle. Universitat Ramon Llull
Passeig Bonanova, 8 08022 Barcelona, Catalonia (Spain)
e-mail: {elisabet,mariasal}@salleURL.edu

Abstract

In this work, a novel classification system for the analysis of mammographic microcalcifications (μCa) using Machine Learning techniques is presented. The automatic classification is performed through Case-Based Reasoning (CBR), which integrates in one system two different characteristics: machine learning capabilities and problem solving capabilities. CBR uses a similar philosophy to that which humans sometimes use: it tries to solve new cases (examples) of a problem by using old previously solved cases. We studied the application of CBR classification to the problem of differentiate benign from malignant μCa in mammograms, obtained from the mammography database of the Girona Health Area, and compared the classification results to other classification techniques.

Keywords: Microcalcifications, Digital Mammograms, Image Processing, Shape-based Features, Feature Evaluation, Machine Learning, Case-Based Reasoning, Automatic Classification, Diagnosis, Classifier Systems.

1. Introduction

Breast cancer is the most common cancer of western women and is the leading cause of cancer-related death among women aged 15-54 (Fajardo and Williams 1996). Survival from breast cancer is directly related to the stage at diagnosis. The earlier the detection, the higher chances of successful treatment (Vyborny and Giger 1994). In an attempt to improve early detection, a study has been undertaken to process the screening mammograms of breast cancer patients in order to analyze the microcalcifications features that help to differentiate benign from malignant cases.

A number of CAD (Computer Aided Diagnosis) techniques have been developed for the detection and classification of microcalcifications (Jiang, Nishikawa and Papaioannou 1998; Tsuji, Freedman, and Mun 1999; Veldkamp and Karssemeijer 1996), in which several image processing techniques are applied ranging from gray-level image analysis (Dhawan and Chitre 1996) to morphological methods (Mossi and Albiol 1998), as well as a great number of classifiers ranging from Bayesian classifiers (Karssemeijer 1991) to neural networks (Tsuji, Freedman, and Mun 1999).

In this paper we propose the use of some selected shape-based features in order to classify clustered microcalcifications between benign and malignant. The computerized analysis of mammographic microcalcifications performed in this work can be divided into four steps: 1) digitization of mammograms and enhancement of images; 2) detection and localization of suspicious areas; 3) extraction of shape-based features for every segmented microcalcification in the digitized mammogram; and 4) analysis of the features using Case-Based Reasoning techniques.

The paper is organized as follows: the next section briefly describes the digitization of mammograms and the enhancement of images as well as the detection and localization of suspicious areas, by giving an overview of the acquisition and segmentation procedures; section 3 summarizes the process of CBR analysis applied to the shape features selected for the μCa. Some experimental results are given in section 4, comparing to previous results using statistical methods. Finally, the paper ends with some conclusions and suggest promising directions.

2. Materials and methods

The study is composed by two separate stages: a retrospective and a prospective one. Each mammogram contains one or more clusters of suspicious microcalcifications.
A set of 146 mammograms was used at the retrospective stage with the goal to analyse the incidence of features in the malignant character of the microcalcifications and therefore, to choose the best features in order to build a statistical predictive model for the malignant diagnosis. The real diagnosis was known in advance from biopsies.
The medical diagnosis for the retrospective mammograms, issued by expert radiologists and oncologists, are known.
The predictive model was tested at a prospective stage, composed by 70 mammograms not diagnosed in advance. In order to evaluate the performance of the selected features for characterizing the microcalcifications and the power of the statistical model, the diagnosis provided by the model was compared to the real diagnosis given by the biopsies. Finally, this evaluation was compared to the diagnosis issued by 3 expert radiologists.

2.1 Image Digitization

Conventional mammograms, in which the positions of clustered microcalcifications were determined by well experienced radiologists, were digitized using a CCD camera at a pixel size ranging from 12 to 37 micrometers and a twelve-bit gray scope, producing a 1524×1012 matrix image. An unshap-mask filter was applied to enhance the high-frequency component on the digitized images, only to make it easier for the observers to recognize the microcalcifications at the stage of annotated image display.

The whole set of digitized mammograms composes an unpublished database formed by patients of the Regional Health Area of Girona, now available upon request, which in the future may contribute to increase the digital mammogram databases.

2.2 Image Segmentation

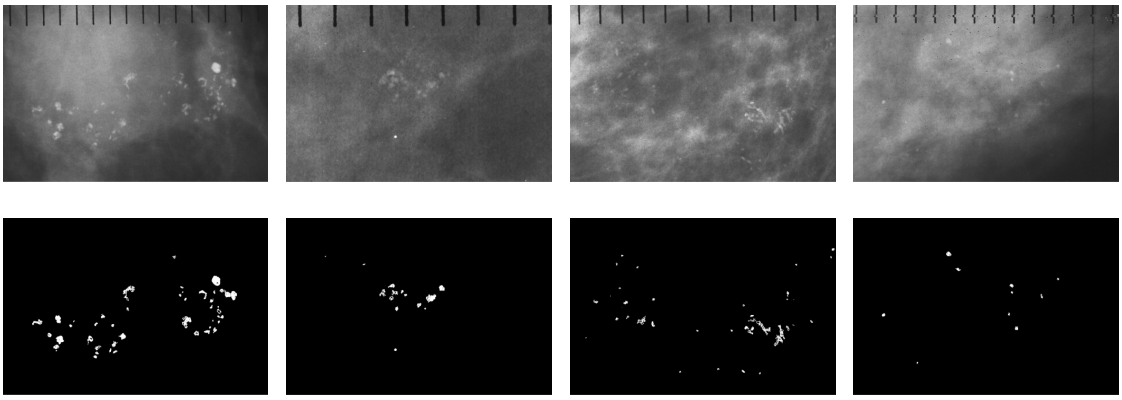
The microcalcifications are segmented using a region-growing algorithm based on *Shen* segmentation techniques (Shen, Rangayyan, and Dessautels 1994). The algorithm starts with a selected pixel inside every microcalcification, called the seed pixel, which has been manually selected by the expert radiologists. This becomes the first region pixel; then, pixels *p(i,j)* of every 4-connected neighbor of pixels belonging to the region, are checked for the tolerance-region condition:

$$(1+\tau)(F_{max}+F_{min})/2 \geq p(i,j) \geq (1-\tau)(F_{max}+F_{min0})/2 \quad (1)$$

where F_{max} and F_{min} are the current maximum and minimum pixel values of the growing region, and τ is the growth tolerance ($0 \leq \tau \leq 1$). This recursive procedure is continued until no connected pixel meets the condition expressed in the equation (1). Stating that the major difficulty of this method is the determination of the tolerance value for each calcification, a multi-resolution procedure is used, trying to find the most appropriate tolerance value τ for each microcalcification.

A tolerance value τ is selected for each region. The final chosen value is selected among the candidates ranging from 0.01 to 0.4 with a step increment (SS) determined by the seed pixel (SP) value as $SS = 1/SP$. For every region obtained at each tolerance level a feature set is calculated, including shape compactness, centre of gravity *(x,y)* coordinates, and size (number of pixels). The normalized distance of this feature set among the successive tolerance levels is computed, and the feature set with the minimum distance is selected as the final set in order to choose the value of τ.

Figure 1 shows four examples of microcalcification segmentation, using our implementation of the *Shen* algorithm: the first row depicts the original image after its digitization, while the second row shows the segmented microcalcification obtained.



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

Figure 1: Four examples of microcalcification segmentation, based on *Shen* algorithm: (a) well defined and big microcalcifications, (b) poor defined and small microcalcifications, (c) with fat tissue, and (d) in a poorly contrasted image

2.3 Feature extraction

After segmenting the microcalcifications in every digitized mammogram, a set of binary regions was obtained in each image. The characterization of these regions is not a trivial task, although several methods have been proposed in the literature (Woods et al. 1994). Taking into account that shapes and sizes of clustered microcalcifications have been associated with a high risk of carcinoma based on different subjective measures,
such as whether or not the calcifications are irregular, linear, vermiform, branched, rounded or ring like, our efforts were addressed to obtain a feature set related to the shape.

2.4 Initial Feature Set

The shape features initially choosen for characterizing the binary segmented individual μCa are shown in table 1, where a summary description is provided for every feature.

Table 1: Initial feature set used to characterize the segmented microcalcifications

Feature	Description
Area	The number of pixels in the μCa
Perimeter	The total length of boundaries of the μCa
Compactness	Derived from the perimeter (P) and area (A), it is equal to $\frac{P^2}{4\pi A}$
Box Min. X,Y; Max. X,Y	The coordinates of the extreme left, top, right, and bottom pixels, respectively, of the μCa
Feret X,Y	The dimensions of the minimum bounding box of the μCa in the horizontal and vertical directions, respectively
Feret Minimum Diameter	The smallest Feret diameter found after checking a certain number of angles (maximum 64)
Feret Maximum Diameter	The largest Feret diameter found after checking a certain number of angles
Feret Mean Diameter	The average Feret diameter at all angles checked
Feret Elongation	A measure of the shape of the μCa, it is equal to $\frac{FeretMaxDiameter}{FeretMinDiameter}$
Number of Holes	The number of holes in the μCa
Convex Perimeter	An approximation of the perimeter of the convex hull of the μCa
Roughness	A measure of the roughness, it is equal to $\frac{Perimeter}{ConvexPerimeter}$
Length	A measure of the true length of the μCa
Breadth	A measure of the true breadth
Elongation	Equal to $\frac{Length}{Breadth}$
Centroid X,Y	The *(x,y)* position of the center of gravity of the μCa

We must remind that the previous features concern every segmented microcalcification in the mammogram, so it was necessary to modify the features in order to stablish reliable comparisons among mammograms.

3. Case-Based Reasoning for feature selection

Case-Based Reasoning (CBR) integrates in one system two different characteristics: machine learning capabilities and problem solving capabilities. CBR uses a similar philosophy to that

which humans sometimes use: it tries to solve new cases (examples) of a problem by using old previously solved cases (Riesbeck and Schank 1989). The process of solving new cases contributes with new information and new knowledge to the system. This new information can be used for solving other future cases. The basic method can be easily described in terms of its four phases (Aamodt and Plaza 1994). The first phase *retrieves* old solved cases similar to the new one. In the second phase, the system tries to *reuse* the solutions of the previously retrieved cases for solving the new case. The third phase *revises* the proposed solution.
Finally, the fourth phase *retains* the useful information obtained when solving the new case. In a Case-Based Classifier system, it is possible to simplify the reuse phase. Reuse can be done by classifying the new case with the same class as the most similar retrieved case.

3.1 CaB-CS and extensions

We use the CaB-CS (Case-Based Classifier System) (Garrell et al. 1999; Llorà et al. 2000) and some extensions (Salamó et al. 1999; Llorà et al. 2000). CaB-CS allows the user to test several variants of CBR.
To be exact, the variants presented in this paper are focused on the *retrieval* phase (phase 1). Phase 1 retrieves the *most similar* case or cases to the new case. Obviously, the meaning of *most similar* will be a key concept in the whole system. Similarity between two cases is computed using different similarity measures.
For the problem that we present in this paper, we use the principal similarity functions of the CaB-CS, more same extensions. To be exact, the different similarity functions can be classified in two groups: 1) Similarity functions based on the distance concept: Minkowski's metric (Hamming, Euclidean and Cubic distance), Clark distance, and Cosine distance; and 2) Similarity functions based on spheres: Sphere of proximity, Sphere MinMax and Sphere Mean (these functions were proposed by Golobardes (Llorà et al. 2000)).

3.2 Similarity functions based on the distance

The most used similarity function is the *Nearest Neighbour (NN)* algorithm, which computes the similarity between two cases using a global similarity measure. The practical implementation (used in CaB-CS) of this function is based on the *Minkowski's metric*, and some extensions of CaB-CS includes the *Clark distance* and the *Cosine distance*.

Minkowski's metric. The Minkowski's metric performs the similarity between two cases, computing the sum of all the distance-weighted difference between both feature values.
In this study we test the Minkowski's metric for three different variants: *Hamming distance*, *Euclidean distance*, and *Cubic distance*.

Clark's Distance. The Clark's distance is defined like the Minkowski's metric but each individual feature differences are balanced by the sum of each feature values.

Cosine Distance. The Cosine distance is based on vector properties in an Euclidean space. It measures the Cosine angle in an n-dimensional vector space.

3.3 Similarity functions based on spheres

CaB-CS proposes other similarity functions based on the *sphere* concepts (Llorà et al. 2000). These functions search some sphere able to *explain* the new case -that we want to solve-. The first and the second function proposed, the *Sphere of Proximity* and the *MinMax Sphere*,

compute the similarity between two cases using a local similarity measure, but the third function, the *Mean Sphere*, computes the similarity using a global similarity measure.

Sphere of Proximity. The Sphere of Proximity searches cases from the case memory that are into a delimited sphere that describes the new case, feature by feature. So, we say that two cases are similar if they are also similar feature by feature. The sphere boundaries are computed using the variance -of the class which belongs to the retrieval case- for each feature. In this sense, we select the cases from the case memory if they satisfy the following condition:

$$\text{If } \forall a_i : \Delta^2 a_i \leq threshold \times \left(\frac{Variance(a_i)}{iteration} \right) \text{ then}$$
$$\text{include this case in the } list\_of\_selected\_cases; \tag{2}$$

Where a_i is the *ith* feature; $\Delta^2 a_i$ is the squared difference between both values of the *ith* feature -for the new case and the retrieved case-; the *threshold* weighs the relevance of the *ith* feature; the *iteration* represents the number of tries that the function computes in order to obtain a correct classification; the *list_of_selected_cases* is the list where the function retains the "similar" cases; and the *variance* of the feature a_i is computed as:

$$Variance(a_i) = \frac{\sum_{j=1}^{N} (x_{ij} - \overline{x_i})^2}{N-1} \tag{3}$$

Where N is the cardinality of the case memory (the number of cases); x_{ij} is the value of the feature i for the case j; and $\overline{x_i}$ is the mean of the *ith* feature.
If we obtain an empty *list_of_selected_cases* then we can not classify the new case, otherwise we can use different criteria in order to choose the most similar case to the new case.

MinMax Sphere. The MinMax Sphere computes one *sphere* for each class -in which we can classify the new case-. Each sphere -of any class *C*- contains information for each feature about the minimum and maximum values, based on the cases of the case memory that belong to this class. In this sense, this similarity function classifies a new case in the class *C* if, for all features, it satisfies that:

$$\begin{aligned} \forall a_i : (value\_min(C, a_i) \times threshold\_min) \leq \\ value(New\_case, a_i) \leq \\ (value\_max(C, a_i) \times threshold\_max) \end{aligned} \tag{4}$$

Where F is the number of features that describes the case; a_i is the feature i; $value\_min(C, a_i)$ and $value\_max(C, a_i)$ are the minimum value and the maximum value of the sphere of the class C for the *ith* feature; $value(New\_case, a_i)$ represents the value of the *ith* feature of the new case; and the *threshold_min* and the *threshold_max* weighs the relevance of the *ith* feature for the minimum value and the maximum value respectively.

Mean Sphere. The Mean Sphere also computes one *sphere* for each class -in which we can classify the new case-. Each sphere -of any class *C*- contains information for each feature about the mean value based on the cases of the case memory that belongs to this class.

Now, the Mean Sphere function uses a similarity function based on distance (e.g. Hamming distance) in order to retrieve the "most similar sphere" to the new case. In this sense, we say that this function uses a global similarity measure.

3.4 Weights

Minkowski's metric and functions based on spheres need to weigh the feature relevance. In this paper the system works using three different options: *Without weights*, *Sample correlation* and *Shannon's Entropy*.

Without weights. The worst possibility (but it is the more frequent) is no information about the features relevance, and there are not automatic methods that could compute its weights. In this case, we consider that for each weight *i*, its value is *1.0*.

Sample correlation. CaB-CS uses *sample correlation* in order to compute the weights w_i that weigh the relevance of the features *i*. In other words, the weights are performed by the sample correlation which exists between each feature x_i and the class *y*.

Shannon's Entropy. In this way, CaB-CS uses the *information-theoretic entropy measure* in order to compute "the relevance" of the features. The entropy measure tends to direct the search in the direction of more significant features.
In the next section, we present the results obtained by the CaB-CS system.

4. Results

The information used to feed the machine learning systems can be summarized as follows: once the image processing phases are completed, an $m \times 23$ real valued matrix is obtained for each mammograph, which contains as many rows (*m)* as the number of microcalcifications are analyzed in the image, while the number of columns (*23)* are related to the computed shape features for every microcalcification. In order to feed this information to the machine learning system (CaB-CS and extensions), the matrix is flattened into a vector. This process is achieved computing the mean value of each feature of the microcalcifications present in the image. Therefore, an image can be reduced to a real-valued vector with *23* features.
The human experts also decided which *training* and *test* sets must be used. The *training* set contained 146 samples, while the *test* set had 70 samples. The obtained results were evaluated comparing the result of this classification with the diagnosis given by the biopsies.

4.1 Results using statistical methods

In (Martí et al. 1998) a statistical prediction model was developed. This statistical model was based on regression, and a *logit* function was used in order to obtain which features are relevant to the classification process. The results obtained never outperformed the 51% of success that human experts were able to reach.

4.2 Results using CaB-CS

In this subsection we present the results using a CBR approach. In fact, we present the results using the CaB-CS system and their extensions.

The table 2 shows the results using the different similarity functions: Hamming distance, Euclidean distance, Cubic distance, Clark's distance, Cosine distance, Sphere of Proximity, MinMax Sphere and Mean Sphere, and which feature selection method has been used.

Table 2: Results using the CBR approach

Sim. Function	%Correct	%Incorrect	Weighting method
Hamming	72.9	27.1	Without w_i
Euclidian	72.9	27.1	Without w_i & Correlation
Cubic	**74.3**	25.7	Correlation
Clark	**77.1**	**22.8**	**Correlation**
Cosine[1]	64.3	25.7	Without w_i
Proximity	72.9	27.1	Shannon's Entropy
MinMax	72.9	27.1	Without w_i & Correlation
Mean	72.9	27.1	Without w_i & Correlation & Shannon's Entropy

We want to remark that the different similarity functions retrieve the most similar case to the new case from the case memory, using very different policies. On one hand, we use the more classical view: the similarity functions based on the distance. On the other hand, we present the similarity functions based on spheres, which retrieve the most similar case using -again- different criteria. For example, the function Mean Sphere, use the cases of the case memory in order to construct the spheres that represent the different classes, so these spheres do not represent a real case. Although we use very different points of view, we obtain -as table 2 shows - the same results: 72.9% of *Prediction Accuracy* (PA). And punctually, Cubic and Clark's distances reach the 74.3% and 77.1% of PA, respectively. These results show that these different criteria have similar behaviour on this problem.
The Sample Correlation weighs the linearity among one feature and the class that belongs to. For this problem, the system has been obtained a low and very similar linearity for each feature. On the other hand, the Shannon's Entropy tends to search for each feature its discriminate capability. Again, for this problem, the system has been computed a low and very similar discriminate capability for each feature. In this sense, the results among three weigh criteria are very similar.
Also, we want to remark that these results are the best results after trying about 500 different options of the CaB-CS and extensions. But, almost all results are very close.

The classification method was evaluated using the *test* database composed by 70 cases (21 malignant and 49 benign) of which all had been sent to biopsy. The CaB-CS system achieved an ROC area of 0.82 in the task of distinguishing benign from malignant cases when the Clark distance with correlation was used, yielding a sensitivity of 82% at a specificity of 66%.

5. Conclusions

We have presented a novel classification system for the analysis of mammographic microcalcifications. The presented method makes use of a Case-Based Classifier System (CaB-CS), with a similar philosphy to that used by the humans: new cases of problem are solved by using old previously cases. The obtained results suggest the following for future

[1] The Cosine distance diagnoses 64.286% correctly, 25.714% incorrectly, and for a 10% is not clear its diagnostic. So it diagnoses a 71.42% correctly among all the classified cases.

study: incorporating more features to the microcalcification characterization than the shape-based features used, and examining new similarity functions for the evaluation of new cases. Moreover, we are currently incorporating our classification schemes into an intelligent mammography workstation that we are developing for use as a "second opinion" in breast cancer screening, initially intended to assist radiologists in classifying malignant and benign clustered microcalcifications in mammograms.

Acknowledgements
The authors wish to thank the University Hospital "Dr. Josep Trueta" as well as the private and public hospitals of the Regional Health Area of Girona, for providing them with the mammographic images utilized in this work. This work was supported by the *Fondo de Investigación Sanitaria* of Spain, Grant No. 00/0033. The results of this work were obtained using the equipment co-funded by the *Direcció de Recerca de la Generalitat de Catalunya (D.O.G.C 30/12/1997)*.

References
Aamodt, A., and E. Plaza. 1994.Case-based reasoning: Foundations issues, methodological variations, and system approaches. *AI Communications*, 7: 39-59.

Dhawan, A.P., and Y. Chitre. June 1996. Analysis of Mammographic Microcalcifications using Gray-level Image Structure Features. *IEEE Transactions of Medical Imaging*, (15) pp. 246-259.

Fajardo, L.J., and M.B. Williams. 1996. The Clinical Potential of Digital Mammography. In *Proceedings of the 3rd International Workshop on Digital Mammography*, pp. 43-52. Chicago (USA).

Garrell, J.M., E. Golobardes, E. Bernadó, and X. Llorà. October 1999. Automatic diagnosis with genetic algorithms and case-based reasoning. *Artificial Intelligence in Engineering*, 13(4):367-362. Elsevier Science Ltd., ISSN 0954-1810.

Jiang, Y., R.M. Nishikawa and J. Papaioannou. 1998. Requirement of Microcalcification Detection for Computerized Classification of Malignant and Benign Clustered Microcalcifications. In *Proceedings of the SPIE Conference on Image Processing*, vol. 3338, pp. 313-317. San Diego (USA).

Karssemeijer, N. 1991. A Stochastic Model for Automated Detection of Calcifications in Digital Mammography. In *Proceedings of the 12th International Conference on Information Processing in Medical Imaging*, pp. 227-238.

Llorà, X., E. Golobardes, M. Salamó, and J. Martí. To Appear in 2000. Diagnosis of Microcalcifications using Case-Based Reasoning and Genetic Algorithms. In *Proceedings of Engineering of Intelligent Systems (EIS'2000)*.

Martí, J., X. Cufí, J. Regincós, J. Español, J. Pont, and C. Barceló. 1998. Shape-based feature selection for microcalcification evaluation. In *Proceedings of the SPIE Medical Imaging Conference on Image Processing*, vol. 3338, pp. 1215-1224.

Mossi, J.M., and A. Albiol. 1996. Automatic Detection of Clustered Microcalcifications using Morphologic Reconstruction. In *Proceedings of the 4th International Workshop on Digital Mammography*, pp. 475-476. Nijmegen (The Netherlands).

Riesbeck, C.K., and R.C. Schank. 1989. Inside Case-Based Reasoning. *Lawrence Erlbaum Associates*, Hillsdale, NJ, US.

Salamó, M., E. Golobardes, J.M. Garrell, and J. Martí. November 1999. Clasificación de microcalcificaciones usando razonamiento basado en casos. In *III Jornadas de Transferencia Tecnológica de Inteligencia Artificial (TTIA'99)*, Murcia, Spain. ISDN 931170-0-5.

Shen, L., R.M. Rangayyan, and J.L. Dessautels. 1994. Detection and classification of mammographic calcifications. In *State of the Art in Digital Mammographic Image Analysis*, K. Bowyer and S. Astley, eds., pp. 198-212, World Scientific Publishing.

Tsuji, O., M.T. Freedman and S.K. Mun. 1999. Classification of Microcalcifications in Digital Mammograms using Trend-oriented Radial Basis Function Neural Network. In *Pattern Recognition*, (32) pp. 891-903.

Veldkamp, W.J.H., and N. Karssemeijer. 1996. Automatic Classification of Clustered Microcalcifications in Digital Mammography. In *Proceedings of the 3rd International Workshop on Digital Mammography*, pp. 231-234. Chicago (USA).

Vyborny, C.J., and M.L. Giger. 1994. Computer Vision and Artificial Intelligence in Mammography. *American Journal of Roentgenology*, (162) pp. 699-708.

Woods, K.S., J.L. Solka, C.E. Priebe, W.P. Kegelmeyer, C.C. Doss, and K.W. Bowyer. 1994. Comparative Evaluation of Patter Recognition Techniques for Detection of Microcalcifications in Mammography. In *State of the Art in Digital Mammographic Image Analysis,* K. Bowyer and S. Astley, eds., pp. 213-231, World Scientific Publishing.