

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,800

Open access books available

122,000

International authors and editors

135M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.

For more information visit www.intechopen.com



Reducing False Positives in a Computer-Aided Diagnosis Scheme for Detecting Breast Microcalcifications: A Quantitative Study with Generalized Additive Models

Javier Roca-Pardiñas¹, María J. Lado¹, Pablo G. Tahoces²
and Carmen Cadarso Suárez²

¹*University of Vigo*

²*University of Santiago de Compostela
Spain*

1. Introduction

Breast cancer continues to be one of the most usual cancers in the world (Siegel et al., 2011). The primary signs that indicate the presence of breast cancer are masses and microcalcifications. Masses can be defined as three-dimensional structures demonstrating convex outward borders, usually evident on two orthogonal views. Microcalcifications are relevant radiologic signs of irregular shape, varying size, and located in an inhomogeneous background of parenchymal tissues. While individual microcalcifications are not, in most cases, clinically significant, clustered microcalcifications appear in 30%-50% of breast cancers (Murphy & DeSchryver, 1978). Moreover, the distribution of the calcification should be specified as grouped, linear, segmental, regional, or diffuse.

It has been demonstrated that an early diagnosis of breast cancer can dramatically reduce the mortality rates. Mammography continues to be the most effective technique for an early detection of the disease, and it is recommended every 1-2 year for women aged between 40-50 years old, and every year for women over 50 years of age. Furthermore, mammography screening should not only be based on age and family history of breast cancer, but also on breast density, among other factors (Schousboe et al., 2011). In fact, mammographic sensitivity for breast cancer can significantly decrease with increasing breast density (Mandelson et al., 2000).

It also deserves comment that radiologists do not detect all the breast cancers present in the mammograms. In fact, the cancers missed at mammographic screening can be categorized into different groups, such as screening errors; minimal sign present; radiographically occult; or radiographically occult at diagnosis (Van Dijck et al., 1993). To minimize the percentage of missed cancers, an independent double reading of mammograms can be an interesting option for increasing the number of breast cancers that are detected at screening mammography (Duijm et al., 2007).

In the last decades, digital mammography has emerged as a promising technique that offers the possibility of a second-opinion consultation, or computer-aided detection (CAD) schemes to assist radiologists in the detection of radiological features that could point to those different pathologies (Banik et al., 2011; Hupse & Karssemeijer, 2009 ; Lado et al., 2001).

Nowadays, utility of CAD systems has been already demonstrated, and there are several computerized systems dedicated to detection and diagnosis tasks approved by the Food and Drug Administration (FDA), such as Second Look (CADx Medical Systems, Inc) (approved in 2002), MammoReader (Intelligent Systems Software, Inc) (approved in 2002), or the Kodak Mammography CAD Engine (Eastman Kodak Company) (approved in 2004).

It is clear that, in order to automatically detect lesions, it could be very useful to learn from the radiologists' experience, as well as to quantify the different image features employed by the clinicians to perform their diagnosis. Even although a computer system will never reach the specialists knowledge level, its ability to detect and classify abnormalities can be improved analyzing the existing differences between the human observer and the computer (Kuprinski & Nishikawa, 1997). It becomes necessary to understand both the medical image contents and the process developed by radiologists for analyzing the information. Given the difficult task of interpreting mammograms by radiologists, the CAD mammographic systems are addressed to limited goals, such as the detection and classification of masses and microcalcifications.

It must be indicated that CAD systems, dedicated to detect abnormalities not only in the breast but also in other medical fields (Doi, 2007), produce suspicious areas that should be identified as lesions or false detections, in order to avoid confusing the clinicians when analyzing the areas detected by the computer. Because of this, a significant stage in nearly all the CAD schemes consists in reducing the number of false positives, by the application of different algorithms and diverse statistical methods(Lado et al., 2006; Tourassi et al., 2005).

There are several models, usually employed by the CAD systems in any field to reduce false detections, such as linear discriminant analysis (LDA) (Yoshida et al., 2002), neural networks (Park et al., 2011), or generalized additive models (GAMs) (Lado et al., 2006). However, reduction of false positives can be a difficult task if an inadequate method or algorithm is selected, this leading to incorrect results, by rejecting correct detections while keeping false positives. Because of this, researchers should pay much attention to the reduction of false positives step.

One of the most important aspects to be considered when the diagnostic imaging systems are analyzed is the evaluation of their diagnostic performance. To perform this task, receiver operating characteristic (ROC) curves are the method usually selected, since they indicate the trade-off between sensitivity and specificity, available from a diagnostic system describing the inherent discrimination capability of these systems (Metz, 1986).

The method of ROC curves can be generalized for the diagnostic performance of both the human observers and the CAD systems. In fact, a large amount of automated systems dedicated to the early detection and diagnosis cancer are frequently evaluated employing ROC methodology, not only in the field of breast lesions (Obuchowski, 2005), but also in nearly any type of cancer or disease (Keotan et al., 2002; Li et al., 2005).

In previous works (Lado et al., 2006; 2008) GAMs were applied to the reduction of false positives in CAD systems dedicated to the detection of microcalcifications. In the first work (Lado et al., 2006), the main goal was to overcome the limitations imposed by LDA in the type

of variable participating in the reduction of false positives. To perform this task, nonlinear classifiers were used, and the methodology was evaluated employing empirical ROC. Results yielded an improvement in sensitivity close to 3%, while the average number of false positive detections was reduced in 0.5 per image.

One of the limitations present in this previous work was that no factors were used in the study. Factors can be defined as categorical variables, such as, for example, type of breast tissue (fatty or dense), than can clearly affect the diagnosis of clusters of microcalcifications, as stated before (Mandelson et al., 2000), and should be taken in consideration, because the response of a continuous covariate may vary across groups defined by levels of a given factor. This indicates that the continuous covariates can behave different in absence/presence of several factors, this producing the corresponding factor-by-curve effects.

To overcome the limitations imposed by the absence of factors in GAMs, the second work (Lado et al., 2008) introduced in the analysis factors and their interactions with continuous variables in the reduction step. The results obtained showed an increase in the sensitivity from more than 2%, while the false positive rate was drastically reduced to the half.

In this work, we propose a new approach to reduce false clustered microcalcifications, employing GAMs and GLMs, which is based on the extraction of several features from the detected clusters, corresponding to both fatty and dense mammograms, and the automated study to discover different behaviours and influences among the covariates (microcalcification features) present in the analysis.

The software programs employed to perform the analysis were developed using R (<http://www.r-project.org/>), an open source software idiom for statistical computing and graphics, which is being used by an increasing number of researchers. Moreover, the R language is distributed under the GNU project, and can run on a wide variety of UNIX, Windows and MacOSX platforms. It is mainly characterized by its core functionality and its high extensibility via the packages, which can be easily downloaded and installed from the CRAN family of Internet sites.

Results show an increase of the sensitivity of the automated system, this leading to a better diagnosis of the disease, not confusing the radiologists by indicating normal areas as suspicious regions, thus reducing the number of biopsies to be performed.

The paper is organized as follows: Section 1 gives a detailed introduction about the breast cancer problem, the automated detection employing CAD systems, and the limitations in detection derived from the use of several features, as well as several solutions and methods employed to perform this task; Section 2 gives an overview about the GLMs and GAMs and the interactions among variables; Section 3 presents the database employed, as well as the CAD system developed for detecting microcalcifications. The database of selected features and the study employing GAMS are presented in Section 4. Section 5 shows and discusses the results obtained. Finally, Section 6 provides the main conclusions of the work. An Appendix also presents the source code developed in R language for performing the GAM analysis.

2. Generalized additive models

In this work, we are interested in predicting the presence or absence of a lesion, using a regression model for binary response. Explicitly, let Y be a binary (0/1) response variable, and $\mathbf{X} = (X_1, \dots, X_q)$ the q -vector of the associated continuous covariates. In this framework,

denoting by $p(\mathbf{X}) = p(Y = 1|\mathbf{X})$, the logistic generalized linear models (GLM) (McCullagh & Nelder, 1989) takes the form:

$$p(\mathbf{X}) = p(Y = 1|\mathbf{X}) = \frac{\exp(a_0 + a_1 \cdot X_1 + \dots + a_q \cdot X_q)}{1 + \exp(a_0 + a_1 \cdot X_1 + \dots + a_q \cdot X_q)} \quad (1)$$

where (a_0, a_1, \dots, a_q) is a vector of coefficients. In some instances, GLMs can be very restrictive, since they assume linearity in the covariates. This constraint can be avoided by replacing the linear index $\eta = a_0 + a_1 \cdot X_1 + \dots + a_q \cdot X_q$ with a non-parametric structure. Accordingly, here we shall concentrate on the generalized additive model (GAM) (Hastie & Tibshirani, 1990), which is a generalization of the GLM, by introducing one-dimensional, non-parametric functions instead of linear components. Specifically, GAMs express the conditional mean

$$p(\mathbf{X}) = p(Y = 1|\mathbf{X}) = \frac{\exp(a + f_1(X_1) + \dots + f_p(X_q))}{1 + \exp(a + f_1(X_1) + \dots + f_p(X_q))} \quad (2)$$

where a is a constant and f_j is the unknown smooth partial function or effect curve associated to each continuous covariate X_j . Note that identification is guaranteed by introducing a constant a into the model and requiring a zero mean for the partial functions. The GAM is widely used as an extension of the traditional GLMs (McCullagh & Nelder, 1989) specially when continuous covariates are present. The GAM is more flexible than the GLM, since the researcher does not assume a parametric form for the effects of the continuous covariates, but only assumes that these effects may be represented by arbitrary unknown smooth functions. The GAMs are easy to interpret, because the additive components simply describe the influence of each covariate separately. Several contributions to GAMs can be found in the literature. Hastie and Tibshirani discussed various approaches using smoothing splines (Hastie & Tibshirani, 1990). Wood introduced a numerical procedure based on regression splines (Wood, 2006). Nowadays, there exists standard software, such as the `mgcv` package in `R`, to fit this model.

A generalization of the "pure" GAM in (2) is the GAM with "factor-by-curve" interactions. In this type of model, the relationship between Y and each of the continuous covariates X_j may vary among the subsets defined by the levels $1, \dots, L$ of a categorical covariate Z (called factor). Explicitly, in the factor-by-curve logistic GAM the effect of each covariate X_j can be expressed as

$$f_j(Z, x) = \begin{cases} f_j^1(x) & \text{if } Z = 1 \\ \vdots & \\ f_j^M(x) & \text{if } Z = M \end{cases}$$

In this way, the effect of each continuous covariate X_j is decomposed in the effects f_j^l associated to each level l ($1, \dots, L$) of the factor Z .

3. Database and CAD system

3.1 Mammogram selection

The mammogram database was constituted by 174, mammograms containing 77 clusters of microcalcifications, proven by biopsy, each mammogram having no more than one cluster.

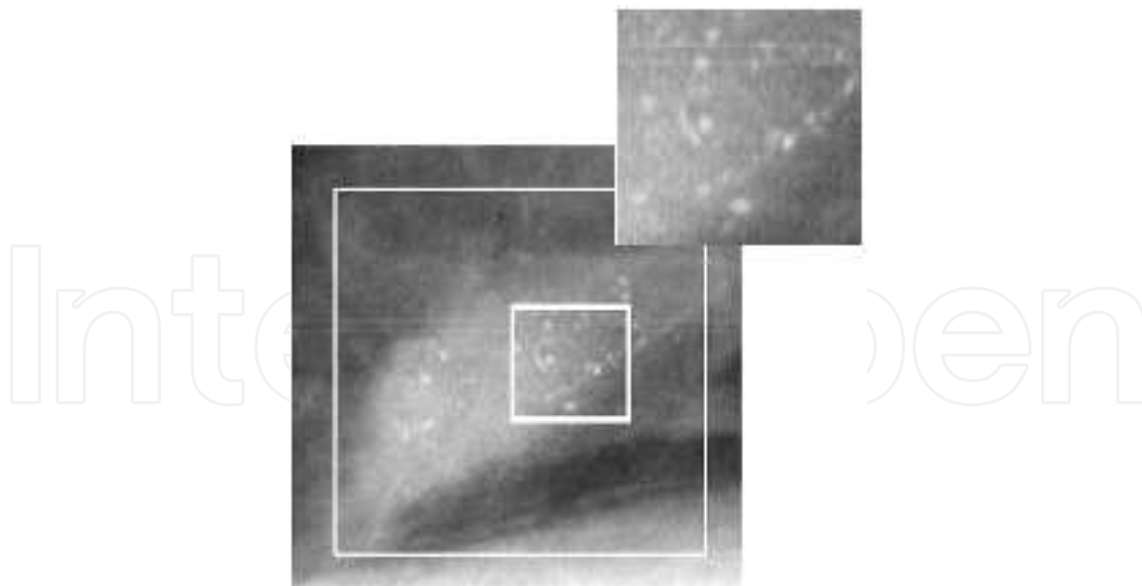


Fig. 1. Region of a mammogram containing a cluster of microcalcifications, delimited by the white square, and a zoomed window containing some microcalcifications

The cases were randomly selected from the mammographic screening program, currently underway, from 1992, at the Galicia Community (Spain). This program is integrated in the European Network of Reference Centers for Breast Cancer Screening.

The average radiation dose employed for the craniocaudal projections was 1.26 mGy, and 1.49 mGy for mediolateral oblique projection. The radiological classification criteria followed the guidelines stated by the Breast Imaging Reporting and Data System (BI-RADS), which establishes the following groups: a) category 0: need additional imaging evaluation; b) category 1: negative; c) category 2: benign finding, noncancerous; d) category 3: probably benign finding, short-interval follow-up suggested; e) category 4: suspicious abnormality, biopsy considered; f) category 5: highly suggestive of malignancy, appropriate action needed.

All the images were digitized at a resolution of 2000x2500 pixels and 4096 gray levels employing a Lumiscan 85 laser scanner (Lumysis Inc., Sunnyvale, CA).

Two experienced radiologists, by consensus, categorized the mammograms into two groups, according to the breast tissue, resulting in 118 dense mammograms, the rest (56) being classified as fatty mammograms. These two experts also marked the location of each cluster of microcalcification in the digital images, being this marks stored on truth data files, in terms of x and y directions. These data truth archives were used to compare the experimental results, obtained with the use of the computerized system to detect microcalcifications, with the true position of the clusters. Figure 1 shows a region of a mammogram containing clustered microcalcifications.

3.2 CAD system

To detect the clusters of microcalcifications, a CAD system was developed and extensively described elsewhere (Lado et al., 2001). Briefly, the method is a five-step process that includes (Figure 2): a) detection of the breast border, employing a tracking algorithm that computes the gradient of gray levels; b) application of wavelet transform to enhance

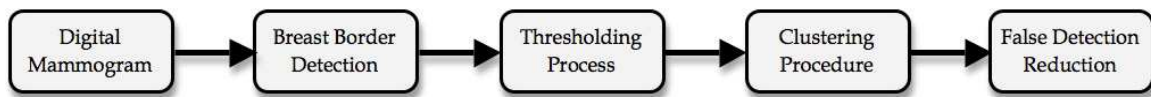


Fig. 2. Scheme of the CAD system for detecting microcalcifications

microcalcifications, by dividing each mammogram into vertical lines, and applying wavelet transform to each lines. As a result, and after applying a local threshold to the wavelet image, a binary image containing the possible seed (origin) points of microcalcifications was obtained; c) gray level thresholding to extract the possible microcalcifications, and application of contrast-size test and morphologic operators, including a region growing algorithm “to grow” the microcalcifications from the corresponding seed points; d) clustering procedure to group the microcalcifications, following the criteria given by Kopans (Kopans, 1989), that considers a cluster of microcalcifications as five or more signals within a region of 1 cm² of area; and 5) reduction of false positives, employing different techniques.

4. Feature extraction and GAM study

When the CAD system previously described was applied over the complete dataset of fatty and dense mammograms, 72 true positives (TPs) were detected, but the system yielded 740 false detections or positives (FP). This means a sensitivity of 93.5% (72/77) at a false positive rate of 4.25 FPs/image.

At this moment, even although the sensitivity produced by the automated system is really promising, it is needed to understand the importance of maintaining a reduced number of false detections, in order to not confuse the radiologist by suggesting normal areas as suspicious, and to reduce the number of biopsies to be performed. Our system arose a high number (4.25) of false detections per image. Because of this, a FP reduction step becomes necessary and fundamental.

To reduce false detections, various features of the detected clusters (true and false positives) were extracted:

1. avglbreast (X1): average gray level value of the breast image containing the detected region, ranging from 0 to 4095 (mean value of 2765.08 ± 275.97).
2. avgIROI (X2): average gray level value of the region of interest (ROI) containing the detected region, ranging from 0 to 4095 (mean value of 2976.70 ± 359.61).
3. avgcluster (X3): average gray level value of the pixels belonging to the detected microcalcifications, ranging from 0 to 4095 (mean value of 3080.32 ± 340.91).
4. avgldist (X4): average distance among the detected microcalcifications in each cluster, measured in pixels (mean value of 74.69 ± 34.63).
5. dimx and dimy (X5 and X6): x and y dimensions of the ROI containing the detected cluster (mean values of 85.25 ± 70.76 and 82.91 ± 69.48 , respectively).
6. size (X7): size of each detected cluster, in pixels (mean value of 10356.02 ± 24558.54).
7. size/avgldist (X8): relationship between the size and the mean distance among microcalcifications for a cluster mean value of 99.26 ± 133.51).
8. size/avgcluster (X9): relationship between the size and the distribution of gray level values of a cluster (mean size of 3.34 ± 7.56).

9. dif1 (X10): difference in gray level values between the average gray level value of the cluster, $avgI_{cluster}$, and the average gray level value of the ROI, $avgI_{ROI}$ (mean size of 103.63 ± 73.16).
10. dif2 (X11): difference in gray level values between the average gray level value of the cluster, $avgI_{cluster}$, and the average gray level value of the breast image, $avgI_{breast}$ (mean size of 315.25 ± 300.00).

The response of the model was the binary variable $true$ (0/1): a value of 0 indicates that the detected cluster is a false positive. If $true$ equals 1, the detected cluster corresponds to a real cluster, and it is a correct detection.

The analysis of the previous feature values was performed employing GAMs, and considering as the factor added to the model the breast tissue (BT), corresponding either to dense tissue (d) or to fatty tissue (f), as previously classified by the radiologists. Explicitly we have considered the following GAM:

$$p(BT, \mathbf{X}) = p(\text{true} = 1 | \mathbf{X}) = \frac{\exp(a + f_1(BT, X_1) + \dots + f_{11}(BT, X_{11}))}{1 + \exp(a + f_1(BT, X_1) + \dots + f_{11}(BT, X_{11}))} \quad (3)$$

with $f_j(BT, x) = f_j^d(x)$ for dense tissue and $f_j(BT, x) = f_j^f(x)$ for fatty tissue.

As stated before, the present work was an attempt to improve the sensitivity of our computerized system by trying the discriminatory capability of different subsets of covariates. A question that tends to arise in regression models of type (3) is that of determining the best subset or subsets of q ($q < 11$) predictors, which will establish the model or models with the best discrimination capacity. As a general rule, as an increasing number of variables are added to the model, the "apparent" fit of the observed data will be improved. However, these estimates are not always satisfied, due to various reasons. On the one hand, inclusion of irrelevant variables would increase the variance of the estimates, thereby amounting to a loss of the predictive capacity of the model; and on the other, inclusion of many variables would mean that the model would be difficult to interpret.

To choose the model, we have used an automatic forward stepwise selection procedure. This procedure selects the model containing the best subset of q variables that would provide the best discrimination capacity, and eliminates the remainder from the model, according to an optimal criterion based on the use of the ROC curve. The area under the ROC curve (AUC) is one of the most widely used criteria for comparing the performance of a series of binary response regression models. The ROC curve relies on false/true-positive/negative tests, where sensitivity is the proportion of event responses that were predicted to be events and specificity is the proportion of non-event responses that were predicted to be non-events. The plot of sensitivity (i.e., hit rate) versus 1-specificity (i.e., false alarm rate) is the ROC curve; the area under this curve measures the accuracy of the detection system and does not require any assumptions concerning either the shape or form of the underlying signal and noise distributions (Saveland & Neuenschwander, 1990). This statistic is a threshold-independent measure of model discrimination, where 0.5 suggests no discrimination, 0.7-0.8 suggests acceptable discrimination, and 0.8-0.9 suggests excellent discrimination (Hosmer & Lemeshow, 2000).

To obtain the corresponding AUCs for various and different covariate subsets, the models (3) were trained on half of the outputs of the detection scheme, which resulted in 36 true

positives and 370 false detections, corresponding to 90 mammograms. The cases employed for training the technique were randomly extracted from the total number of outputs of the CAD scheme. The models were finally tested on the other half of the cases: 36 true positives and 370 false positives (84 mammograms) that had not been used at the initial training stage. The performances of the developed GAMs and GLMs, with the different feature values, were analyzed employing ROC analysis, and considering as the decision variable the estimated probabilities obtained with the models.

To obtain the corresponding AUCs for various and different covariate subsets, the models (3) were trained on half of the outputs of the detection scheme, randomly extracted from the total number of outputs of the CAD system. The models were finally tested on the other half of the cases that had not been used at the initial training stage. The performances of the developed GAMs and GLMs, with the different feature values, were analyzed employing ROC analysis, and considering as the decision variable the estimated probabilities obtained with the models.

5. Results and discussion

This research work aimed at studying how the different features extracted from true and false positive clusters of microcalcifications behave in presence of categorical covariates and factors that can influence and even condition their behaviour. The main goal is, in this way, to discriminate between true clusters and false detections.

The interactions among the different variables were considered in the study. Previously to the selection of variables, correlation among the different covariates was calculated in Table (1).

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11
X1	100	56	54	-1	-3	-4	-21	-30	-1	-4	-4
X2	56	100	98	-2	0	-2	-35	60	2	1	-3
X3	54	98	100	1	4	4	-16	64	6	6	2
X4	-1	-2	1	100	74	75	19	3	65	64	66
X5	-3	0	4	74	100	67	20	7	82	88	83
X6	-4	-2	4	75	67	100	25	8	83	86	84
X7	-21	-35	-16	19	20	25	100	1	22	25	23
X8	-30	60	64	3	7	8	1	100	9	10	6
X9	-1	2	6	65	82	83	22	9	100	93	100
X10	-4	1	6	64	88	86	25	10	93	100	94
X11	-4	-3	2	66	83	84	23	6	100	94	100

Table 1. Matrix correlations($\times 100$) between covariates

A high correlation can be observed for several features, particularly between X2 and X3, or among X9, X10 and X11. Surely, this is due to the fact that these variables can be very similar. For example, X2 and X3 represent gray level values for the cluster of microcalcifications and the ROI containing it, and both regions may nearly contain the same pixel values, this resulting in a high similarity between them. However, there are other features with a low correlation, for example the properties based on gray level values and the properties based on either distances or sizes.

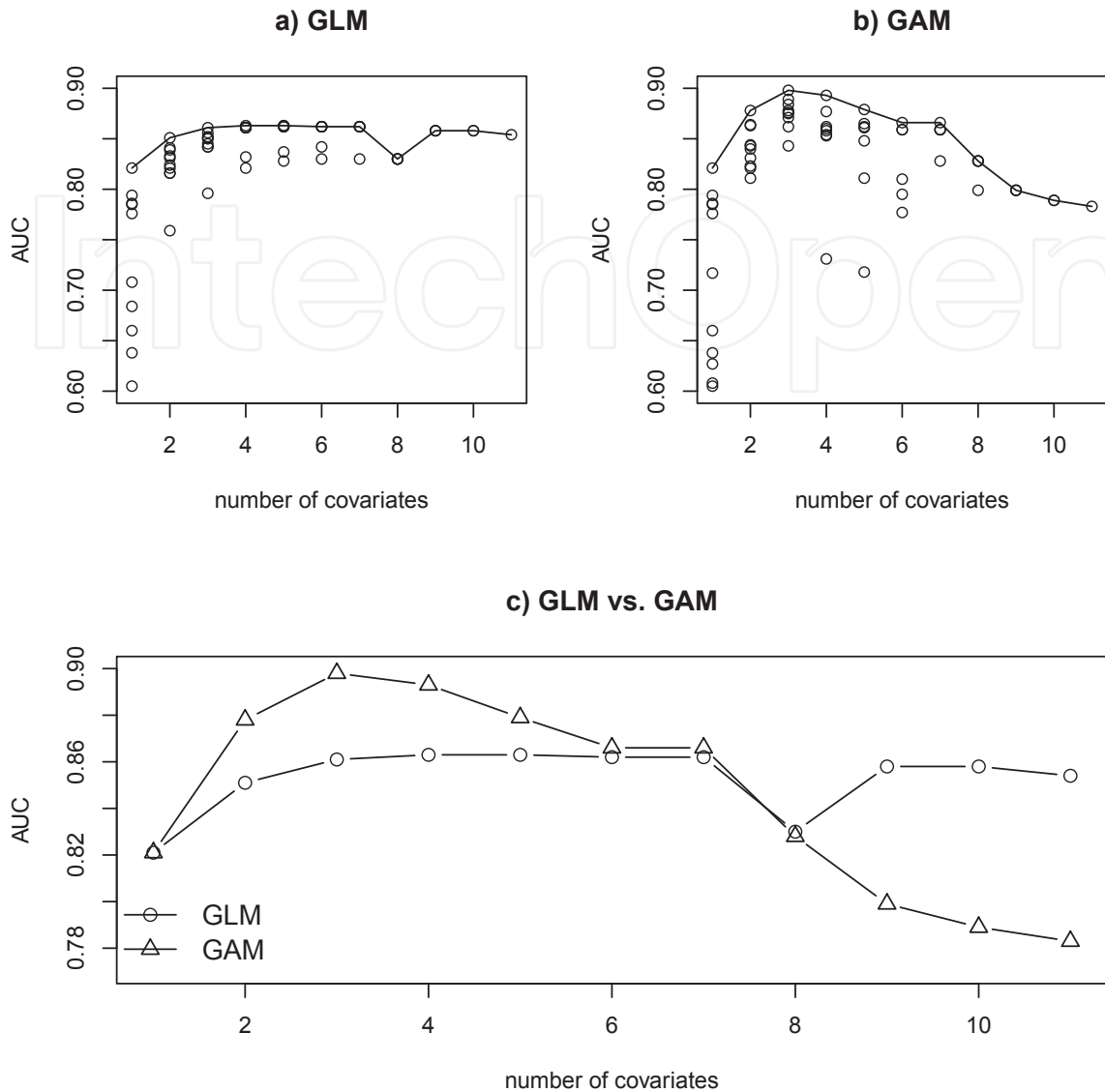


Fig. 3. Possible subset model. For each subset size, the AUC is shown for a) GLM, b) GAM and c) GLM vs. GAM.

When the correlations were calculated, the subset selection for the whole set of covariates was performed, and the corresponding AUCs were obtained for each model. Figure (3) presents the values for all the possible subset model combinations, employing GLMs and GAMs. This figure can be interpreted as follows: The x axis gives the number of covariates included in the statistical model, while the y axis represents the AUC obtained with the model employing the number of covariates indicated by the x value. For example, for a number of covariates equal to 1, a line of vertical points corresponding to different AUC values is represented, the first of them with a value close to 0.60 in both the GLM and the GAM. The last value obtained is greater than 0.80 in both models. All of them correspond to the AUC values obtained with one different covariate between X_1 and X_{11} . The intermediate values are the AUCs yielded by the models constructed employing the rest of covariates.

From Figure (3), it can be observed that, if only one covariate is considered, the best AUC obtained in the GLM is lower than in the rest of cases; however, the AUC is very similar when selecting 2 or more covariates.

When applying the GAM the situation is different, providing the best results for the AUC values when three covariates are considered. Finally, the comparison of the best models for both GLM and GAM indicates that GAM performs better when a number up to 8 covariates are considered. If this number increases, best results are provided by GLMs. Table (2) lists the AUC values obtained for both types of models with different number of covariates.

From Table (2), it can be observed that for both the linear and the additive models, if we only consider one variable in the model, best results were obtained for X10, that is, for the difference in gray level values between the average gray level value of the cluster $avgI_{cluster}$, and the average gray level value of the ROI, $avgI_{ROI}$.

q	GLM		GAM	
	variables	AUC	variables	AUC
1	10	82.10	10	82.10
2	8	85.10	1	87.80
3	4	86.10	11	89.80
4	2	86.30	8	89.30
5	3	86.30	9	87.90
6	1	86.20	3	86.60
7	5	86.20	4	86.60
8	6	83.00	2	82.80
9	7	85.80	5	79.90
10	9	85.80	6	78.90
11	11	85.40	7	78.30

Table 2. AUC values for the different models and number of covariates in each subset.

As an increasing number of variables are included in the study, that is, as different covariates were considered in both models, greater values for AUCs were obtained; particularly, better results were achieved for GAMs when a number up to 8 covariates were included in the model, and better results were obtained for GLMs for 9 or more variables considered. In the previous models, no distinction was considered for the type of tissue, and a unique analysis was performed for both GLM and GAM. However, to study the effect of the previous models in the breast tissue, different analyses were performed for clusters of microcalcifications embedded on both fatty and dense tissue. Different subset combinations were again obtained for both GLMs and GAMs, and the corresponding AUCs were calculated (Tables (3,4)). From these tables, it can be observed that best AUCs are obtained for fatty tissue, while for dense parenchyma are always lower in the corresponding models. This is consistent with the fact that, for fatty tissue, the contrast value, that is, the difference between the microcalcification and the background surrounding it, is greater than for dense tissue. Thus, detection is

q	fatty		dense	
	variables	AUC	variables	AUC
1	10	79.20	8	73.60
2	3	86.40	7	74.50
3	11	86.70	1	74.90
4	4	87.00	2	74.90
5	2	86.40	3	74.90
6	5	86.40	4	74.90
7	6	85.80	11	72.30
8	7	85.70	5	72.10
9	8	85.70	6	69.90
10	1	85.60	9	73.60
11	9	85.60	10	74.90

Table 3. AUC values for the different GLMs and number of covariates in each subset, for both fatty and dense breast tissue

q	fatty		dense	
	variables	AUC	variables	AUC
1	10	79.20	1	79.00
2	3	86.10	11	84.40
3	5	86.90	5	85.10
4	11	85.30	6	86.80
5	1	81.80	7	86.60
6	9	87.90	3	87.00
7	2	81.20	9	87.70
8	4	80.10	10	87.00
9	6	69.10	2	66.70
10	7	87.00	4	64.10
11	8	62.60	8	70.50

Table 4. AUC values for the different GAMs and number of covariates in each subset, for both fatty and dense breast tissue

improved even for radiologists. The dense parenchyma always presents greater difficulties to perform a correct diagnosis.

Appart from this, we can also perceive that, when one covariate is included in the model, best results are obtained by X10 again; however, if more variables are present, selection is not the same in both GLM and GAMs, and it does not match the selection performed when the tissue

type was not separately considered. Moreover, for fatty tissue GLM performs better, while, by contrary, for dense breasts the optimal results are obtained when employing GLMs to select covariates.

Figure (4) shows the AUC values for the best subset model combinations, employing GLMs and GAMs, for both fatty and dense tissue. Figure 5 represents the AUCs for the global analysis, and for both fatty and dense tissue, for the best GLM and GAM.

It can be observed that, for GLM, results obtained for fatty tissue are higher than those obtained for dense tissue. However, when employing the GAM, differences are lower, and a more reduced number of covariates have to be included in the study.

Sensitivity and false positive rates were also calculated for the best GLM and the best GAM. For linear models, results yielded a sensitivity of 88.31%, at a false positive rate of 3.7 FPs per image. For the same sensitivity, the false positive fraction achieved when reducing false

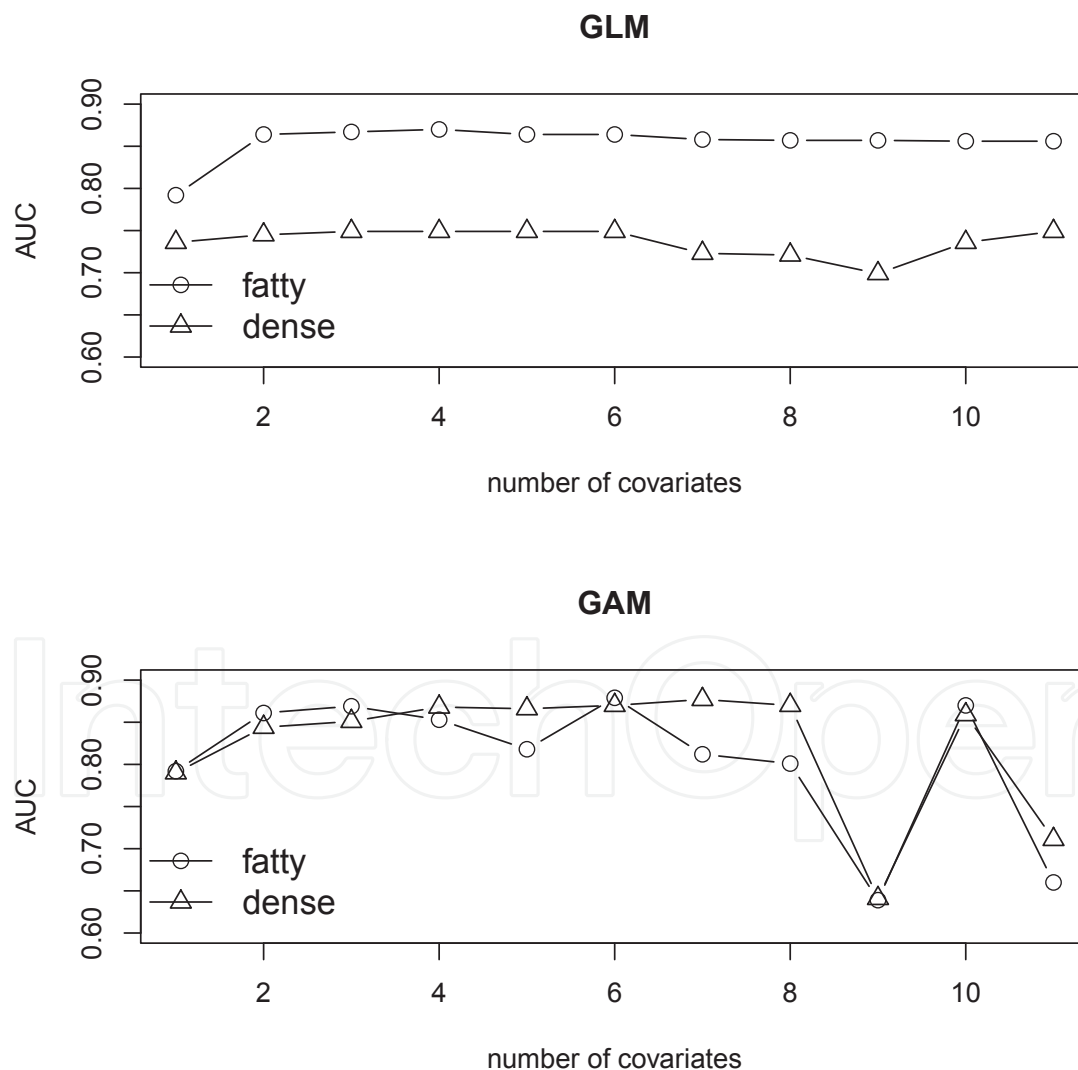


Fig. 4. "Optimal" models for both fatty and dense tissue. For each subset size, the AUC is shown for each model.

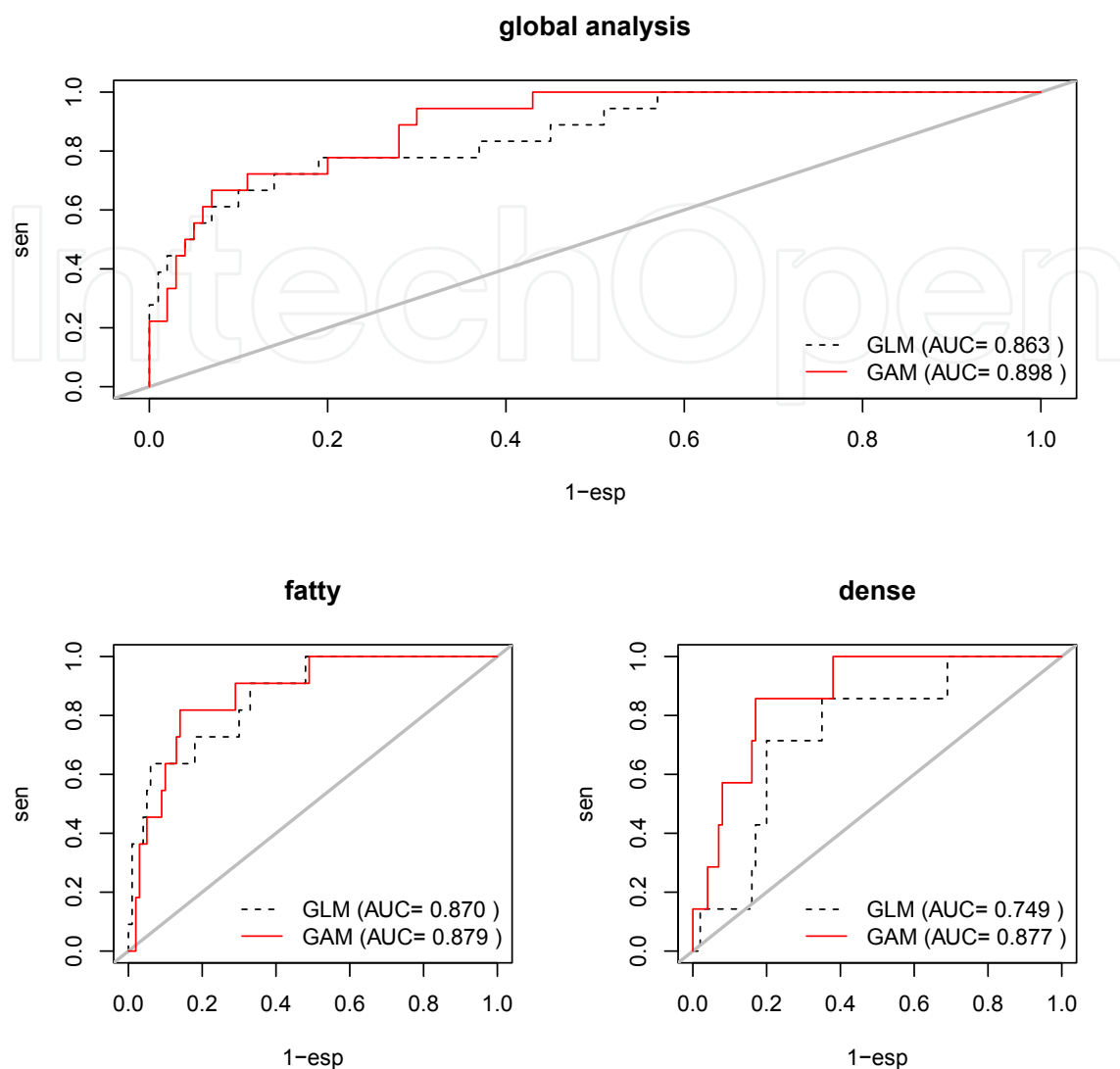


Fig. 5. ROC curves obtained for the GAM and GLM, for the global analysis, and for both types of breast tissue

positives with additive models was 2.7, this demonstrating the benefits of this type of models for discriminating tasks, employing factors and interactions.

6. Conclusion

In this work, GLMs and GAMs were applied to the reduction of false positives yielded by a CAD system devoted to the detection of clusters of microcalcifications. Results indicate that not all the features extracted from the detected clusters are useful for the discrimination between true and false detections: Moreover, there are features that are relevant when the different type of tissue is considered, and their influence is different depending on the breast parenchyma.

After the reduction of false positives, the system is capable of discriminating and detecting clustered microcalcifications from digital mammograms, this suggesting that this CAD

scheme is competent to complement the radiologists' efforts in their daily clinical practice, to help them as second readers in the interpretation of mammograms, and also to improve their diagnostic performance.

Future work will address the issue of reducing the number of false positives, without decreasing the sensitivity, by applying different statistical models to our dataset. Another way to improve the results yielded by the CAD system would be to deal with a higher quality of the digital images, because subtle clusters can be more easily identified in the detection process if the image is digitized at a greater resolution. Thus, to further improve the sensitivity, high quality of the digitized images is required.

7. Acknowledgment

This work has been partially supported by Ministerio de Ciencia e Innovación, MTM2008-01603/MTM. Roca-Pardiñas' research was supported by grant MTM2008-03010 from the Spanish Ministry of Education & Science, and by grants PGIDIT07PXIB300191PR and PGIDIT10PXIB300068 PR from the Galician Regional Authority (Xunta de Galicia).

8. Appendix. Source code developed in R

The R code developed for calculating probability values for the GLMs and GAMs, as well as for obtaining the corresponding AUCs and plots is now given and explained. The starting point of this routines considers data covariates stored on the X covariates, while Y indicates if the data corresponds either to a true detection ($Y = 1$) or to a false positive cluster ($Y = 0$). The factor considered in the study is $F = 0$ for fatty tissue, and $F = 1$ for dense tissue.

The steps followed in the present work were the calculation of the correlation values, and the study of covariates, initially without taking into consideration the type of tissue, calculating in this situation the linear and additive models, employing the functions `stepGLM2` and `stepGAM2` respectively (described in this Appendix), and obtaining for both cases the corresponding areas under the ROC curve, using the `empiricROC` function. Plots and tables were also obtained.

The same analysis was next calculated, but separating the data by the type of tissue where they were embedded in the breast. Graphs, tables and AUC were also obtained. The R code is listed below.

```
#####
# Correlations calculation
cor(X)

#####
# Selection of covariates employing the AUC criterion for the GLMs
# and GAMs

#####
GLMresults=stepGLM(X,Y,"glm"); GAMresults=stepGAM(X,Y)

# Particular studies differentiating the type of tissue
ii<-F==0; X0=X[ii,];Y0=Y[ii]
```

```

ii<-F==1; X1=X[ii,];Y1=Y[ii]
GLM0results=stepGAM(X0,Y0,"glm")
GLM1results=stepGAM(X1,Y1,"glm")
GAM0results=stepGAM(X0,Y0)
GAM1results=stepGAM(X1,Y1)

#####
# Output graphs and tables

# GLMs and GAMs comparison in global study
layout(matrix(c(1,2,3,3), 2, 2, byrow=TRUE))
plot(GLMresults$aucs,xlab="number of covariates",ylab="AUC",
main="GLM",ylim=c(0.6,0.9))
lines(GLMresults$aucopt)

plot(GAMresults$aucs,xlab="number of covariates",ylab="AUC",
main="GAM",ylim=c(0.6,0.9))
lines(GAMresults$aucopt)

plot(GLMresults$aucopt,xlab="number of covariates",ylab="AUC",
ylim=c(0.77,0.90),type='b',pch=1,main="GLM vs. GAM",cex=1.24)
lines(GAMresults$aucopt,type='b',pch=2,cex=1.25)
legend("bottomleft",c("GLM","GAM"),lty=1,pch=c(1,2),
box.lty=0,cex=1.25)

# GLMs and GAMs comparison considering the type of tissue
# GLMs results
par(mfcol=c(2,1))
plot(GLM0results$aucopt,xlab="number of covariates",ylab="AUC",
type='b',pch=1,ylim=c(0.6,0.90),cex=1.25,main="GLM")
lines(GLM1results$aucopt,type='b',pch=2,cex=1.25)
legend("bottomleft",c("fatty","dense"),lty=1,pch=c(2,1),
box.lty=0,cex=1.25)

plot(GAM0results$aucopt,xlab="number of covariates",ylab="AUC",
type='b',pch=1,ylim=c(0.6,0.90),cex=1.25,main="GAM")
lines(GAM1results$aucopt,type='b',pch=2,cex=1.25)
legend("bottomleft",c("fatty","dense"),lty=1,pch=c(2,1),
box.lty=0,cex=1.25)

#####
# Calculation of the GAMs
stepGAM<-function(X,Y,option="gam") {
# X: Covariables matrix
# Y: Response vector
n=dim(X) [1]
nvar=dim(X) [2]

```

```

variables=1:nvar
auc0=NULL; formula0=NULL; inside=NULL
nvars=NULL;aucs=NULL;ROCs=NULL

test<-seq(1,n,4)
Wtraining=rep(1,n);Wtraining[test]=0
for (ivar in 1:nvar) {
  ii=! (variables %in% inside)
  variables=variables[ii]
  auc=NULL
  for (j in 1:length(variables)) {
    if (option=="gam") {formula=paste("+s(X[,",j,"])",sep="")}
else {formula=paste("(X[,",j,"])",sep="")}
    formula=paste(formula0,formula,sep="")
    formula=paste("Y~",formula,sep="")
    if (option=="gam") {modelo=gam(as.formula(formula),
      family="binomial",weights=Wtraining)}
else {modelo=glm(as.formula(formula),family="binomial",
  weights=Wtraining)}
    muhat=predict(modelo,type="response")
    a=EmpiricalROC(Y[Wtraining==0],muhat[Wtraining==0])
    aux=a$auc
    nvars=c(nvars,ivar);aucs=c(aucs,aux);auc=c(auc,aux)
    jj=length(a$t)
    ROCs=rbind(ROCs,cbind(rep(ivar,jj),rep(aux,jj),a$t,a$ROC))}
    inside=c(inside,variables[which.max(auc)])
    auc0=c(auc0,max(auc))
    if (option=="gam") {formula0=paste("s(X[,",inside,"])",
      sep=" ",collapse="+")}
    else { formula0=paste("(X[,",inside,"])",sep=" ",collapse="+")}
  }
return(list(aucs=cbind(nvars,aucs),
aucopt=cbind(1:nvar,auc0),models=inside,roc=ROCs))
}

#####
# Calculation and plotting of the empiric AUCs
EmpiricalROC<-function (group,Y) {
YE=Y[ group==1]; YS=Y[ group==0]
t=seq(0,1,0.01); F=quantile(YS,probs=(1-t),type=1)
ROC=1-sapply(F, function(x) mean( YE <= x ))
ROC=as.vector(ROC)

t=c(0,t,1); ROC=c(0,ROC,1)
m<-length(YS);n<-length(YE)
xmat<-matrix(rep(YS,n),nrow=n,byrow=T)
ymat<-matrix(rep(YE,m),nrow=n,byrow=F)
diffmat<-ymat-xmat
auc<-(length(diffmat[diffmat>0]))+

```

```
0.5*length(diffmat[diffmat==0]))/(m*n)
auc=round(auc,digits=3)
plot(t,ROC,type='s',col='red',xlab="1-esp",ylab="sen")
abline(a=0,b=1,col='grey')
```

```
lines(t,ROC,col='blue',type='s')
text(0.9,0.1,paste("auc=",auc))
result=list(t=t,ROC=ROC, auc=auc)
```

9. References

- Banik, S.; Rangayyan, R.M. & Desautels J.E.L. (2011). Detection of architectural distortion in prior mammograms. *IEEE Transactions on Medical Imaging*, Vol.30, No.2, (February 2011), pp. 279-294, ISSN 0278-0062
- Doi, K. (2007). Computer-aided diagnosis in medical imaging: Historical review, current status and future potential. *Computerized Medical Imaging and Graphics*, Vol.31, No.4, (June 2007), pp. 198-211 ISSN 0895-6111.
- Duijm, L.E.M.; Groenewoud, J.H.; Fracheboud, J. & de Koning, H.J. (2007). Additional double reading of screening mammograms by radiologic technologists: impact on screening performance parameters. *Journal of the National Cancer Institute*, Vol.99, No.15 (August 2007), pp. 1162-1170, ISSN 1460-2105.
- Hastie, T.J. & Tibshirani, J. (1990) *Generalized Additive Models*, Chapman & Hall, ISBN 0-412-34390-8, London.
- Hosmer, D.W. & Lemeshow, S. (2000) *Applied logistic regression*, John Wiley and Sons, ISBN 0-471-35632-8, New York.
- Hupse, R. & Karssemeijer, N. (2009). Use of normal tissue context in computer-aided detection of masses in mammograms. *IEEE Transactions on Medical Imaging*, Vol. 28, No.12, (December 2009), pp. 2033-2041, ISSN 0278-0062
- Keogan, M.T.; Lo, J.Y.; Fred, K.S.; Raptopoulos, V.; Blake, S.; Kamel, I.R.; Weisinger, K.; Rosen, M.P. & Nelson, C.R. (2002). Outcome analysis of patients with acute pancreatitis by using artificial neural network. *Academic Radiology*, Vol. 9, Mo.4, (April 2002), pp. 410-419, ISSN 1076-6332
- Kopans, D.B. (1989). *Breast Imaging*, Lippincott, ISBN 9780397507610, Philadelphia.
- Krupinski, E.A. & Nishikawa, R.M. (1997). Comparison of eye position versus computer identified microcalcification clusters on mammograms. *Medical Physics*, Vol.24, No.1, (January, 1997) pp. 17-23, ISSN 0094-2405
- Lado, M.J.; Tahoces, P.G.; Méndez, A.J.; Souto, M. & Vidal J.J. (2001). Evaluation of an automated wavelet-based system dedicated to the detection of clustered microcalcifications in digital mammograms. *Medical Informatics and the Internet in Medicine*, Vol. 26, No.3, (July 2001), pp. 149-163, ISSN 1463-9238
- Lado, M.J.; Cadarso-Suárez, C.; Roca-Pardiñas, J. & Tahoces, P.G. (2006). Using generalized additive models for construction of non-linear classifiers in computer-aided diagnosis systems. *IEEE Transactions on Information Technology in Biomedicine*, Vol.10, No.4, (April 2006), pp. 246-253, ISSN 1089-7771
- Lado M.J.; Cadarso-Suárez, C.; Roca-Pardiñas, J. & Tahoces, P.G. (2008). Categorical variables, interactions and generalized additive models. Applications in computer-aided diagnosis systems. *Computer in Biology and Medicine*, Vol.38, No.4, (April 2008), pp. 475-483, ISSN 1879-0534

- Li, F.; Arimura, H.; Suzuki, K.; Shiraishi, J.; Li, Q.; Abe, H.; Engelmann, R.; Sone, S.; MacMahon, H. & Doi, K. (2005). Computer-aided detection of peripheral lung cancers missed at CT: ROC analyses without and with localization. *Radiology*, Vol.237, No.11, (November 2005), pp. 684-690, ISSN 1527-1315
- McCullagh, P. & Nelder, J.A. (1989) *Generalized Linear Models*, Chapman & Hall, ISBN 0-412-31760-5, London.
- Mandelson, M.T.; Oestreicher, N.; Porter, P.L.; White, D.; Finder, C.A.; Tapli, S.H. & White, E. (2000). Breast density as a predictor of mammographic detection: comparison of interval- and screen-detected cancers. *Journal of the National Cancer Institute*, Vol.92, No.13 (July 2000), pp. 1081-1087, ISSN 1460-2105
- Metz, C.E. (1986). ROC methodology in radiologic imaging. *Investigative Radiology*, Vol.21, No.9, September(1986), pp. 720-733, ISSN 0020-9996
- Murphy, W.A. & DeSchryver-Kecskemeti, K. (1978). Isolated clustered microcalcifications in the breast: Radiologic-pathologic correlation. *Radiology*, Vol.127, No.5, (May 1978), pp. 335-341, ISSN 1527-1315
- Obuchowski, N.A. (2005). ROC analysis. *American Journal of Roentgenology*, Vol. 184, No.2, (February 2005), pp. 364-372, ISSN 1546-3141
- Park, S.C.; Chapman, B.E. & Zhen, B. (2011). A multistage approach to improve performance of computer-aided detection of pulmonary embolisms depicted on CT images: Preliminary investigation. *IEEE Transactions on Biomedical Engineering*, Vol.58, No.6, (June 2011), pp. 1519-1527, ISSN 0018-9294
- Schousboe, J.T.; Kerlikowske, K.; Loh, A.; Cummings, S.R. (2011). Personalizing mammography by breast density and other risk factors for breast cancer: analysis of health benefits and cost-effectiveness. *Annals of Internal Medicine*, Vol.155, No.1, (July 2011), pp. 10-20, ISSN 1539-3704
- Saveland, J.M. & Neuenschwander, L.F. (1990) A signal detection framework to evaluate models of tree mortality following fire damage. *Forest Science*, Vol.36, No.1, (January 1990), pp. 66-76, ISSN 1420-1143
- Siegel, R.; Ward, E.; Brawley, O. & Jemal, A. (2011). Cancer statistics, 2011. The impact of eliminating socioeconomic and racial disparities on premature cancer deaths. *CA Cancer Journal for Clinicians*, Vol.61, No.4, (August 2011), pp. 212-236, ISSN 1542-4863
- Tourassi, G.D.; Eltonsy, N.H.; Graham, J.H.; Floyd, C.E. & Elmaghraby, A.S. (2005). Feature and knowledge based analysis for reduction of false positives in the computerized detection of masses in screening mammography. *Proceedings of the 27th Annual Conference for the IEEE EMB*, pp. 6524-6527, ISBN 0-7803-8741-4, Shanghai, China, September 1-4, 2005.
- Van Dijck, J.A.; Verbeek, A.L.; Hendriks, J.H. and Holland, R. (1993). The current detectability of breast cancer in a mammographic screening program. A review of the previous mammograms of interval and screen-detected cancers. *Cancer*, Vol.72, No. 6, (June 1993) pp. 1933-1938, ISSN 1097-0142
- Wood, S. (2006) *Generalized additive models: an introduction with R*, CRC Statistics, ISBN 9781584884743, United Kingdom.
- Yoshida, H.; Masutani, Y.; MacEaney, P.; Rubin, D.T. & Dachman, A.H. (2002). Computerized detection of colonic polyps at CT colonography on the basis of volumetric features: Pilot study. *Radiology*, Vol.222, No.2, (February 2002), pp. 327-336, ISSN 1527-1315



Cancer Prevention - From Mechanisms to Translational Benefits

Edited by Dr. Alexandros G. Georgakilas

ISBN 978-953-51-0547-3

Hard cover, 476 pages

Publisher InTech

Published online 20, April, 2012

Published in print edition April, 2012

This unique synthesis of chapters from top experts in their fields targets the unique and significant area of cancer prevention for different types of cancers. Perspective readers are invited to go through novel ideas and current developments in the field of molecular mechanisms for cancer prevention, epidemiological studies, antioxidant therapies and diets, as well as clinical aspects and new advances in prognosis and avoidance of cancer. The primary target audience for the book includes PhD students, researchers, biologists, medical doctors and professionals who are interested in mechanistic studies on cancer prevention and translational benefits for optimized cancer treatment.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Istvan Feher, Javier Roca-Pardinas, Maria J. Lado, Pablo G. Tahoces and Carmen Cadarso Suarez (2012). Reducing False Positives in a Computer-Aided Diagnosis Scheme for Detecting Breast Microcalcifications: A Quantitative Study with Generalized Additive Models, *Cancer Prevention - From Mechanisms to Translational Benefits*, Dr. Alexandros G. Georgakilas (Ed.), ISBN: 978-953-51-0547-3, InTech, Available from: <http://www.intechopen.com/books/cancer-prevention-from-mechanisms-to-translational-benefits/reducing-false-positives-in-a-cad-scheme-for-detecting-breast-microcalcifications-a-quantitative-study>

INTECH
open science | open minds

InTech Europe

University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China

Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2012 The Author(s). Licensee IntechOpen. This is an open access article distributed under the terms of the [Creative Commons Attribution 3.0 License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

IntechOpen

IntechOpen