

# INFLUENCE FUNCTION BASED GAUSSIANTY TESTS FOR DETECTION OF MICROCALCIFICATIONS IN MAMMOGRAM IMAGES

M. Nafi Gürcan<sup>1</sup>, Yasemin Yardımcı<sup>2</sup>, A. Enis Çetin<sup>1</sup>

<sup>1</sup> Department of Electrical and Electronics Engineering, Bilkent University, Ankara, Turkey

<sup>2</sup> Department of Computer Engineering, METU, Ankara, Turkey

E-mail: cetin@ee.bilkent.edu.tr

## ABSTRACT

In this paper, computer-aided diagnosis of microcalcifications in mammogram images is considered. Microcalcification clusters are an early sign of breast cancer. Microcalcifications appear as single bright spots in mammogram images. We propose an effective method for the detection of these abnormalities. The first step of this method is two-dimensional adaptive filtering. The filtering produces an error image which is divided into overlapping square regions. In each square region, a Gaussianity test is applied. Since microcalcifications have an impulsive appearance, they are treated as outliers. In regions with no microcalcifications, the distribution of the error image is almost Gaussian, on the other hand, in regions containing microcalcification clusters, the distribution deviates from Gaussianity. Using the theory of the influence function and sensitivity curves, we develop a Gaussianity test. Microcalcification clusters are detected using the Gaussianity test. Computer simulation studies are presented.

## 1. INTRODUCTION

Breast cancer is one of the most deadly diseases for women. The survival rate approaches 100 percent if cancer is detected early. Microcalcifications are an early sign of breast cancer and they appear as single, bright spots on mammograms (X-ray images of breasts). Because they are small and subtle, microcalcifications are difficult to detect by radiologists. In this work, we develop a computer-aided diagnosis (CAD) scheme for the detection of microcalcification clusters.

Recently, we developed CAD schemes for the computerized detection of microcalcifications based on higher order statistics, adaptive filtering and Gaussianity tests [1, 2, 3]. In these schemes, we make use of two-dimensional (2-D) adaptive filtering and a Gaussianity test recently developed by Ojeda et al. (the OCM test

for short) for causal invertible time series [4].

In our method, a Least Mean Square (LMS) type 2-D adaptive filter is used. The adaptive filter predicts an image pixel  $x[m, n]$  at location  $(m, n)$  as a weighted average of pixels in its region of support. The region of support,  $\mathcal{R}$ , of the filter is chosen as the pixels surrounding the pixel to be predicted. The predicted value  $\hat{x}[m, n]$  is given as

$$\sum_{\substack{k=-n_1 \\ (k,l) \neq (0,0)}}^{n_1} \sum_{l=-n_2}^{n_2} w_{(m,n)}[k,l] x[m-k, n-l],$$

$$m = 0, \dots, N_1 - 1, \quad n = 0, \dots, N_2 - 1 \quad (1)$$

where  $x$  is the input image of size  $N_1 \times N_2$ ,  $w_{(m,n)}$  are the weight values at  $(m, n)$ , and  $(2n_1 + 1) \times (2n_2 + 1)$  is the size of the region of support,  $\mathcal{R}$  of the adaptive filter.

The prediction error at pixel location  $(m, n)$  is computed as the difference between the predicted pixel value,  $\hat{x}[m, n]$ , and the actual pixel value,  $x[m, n]$

$$e[m, n] = \hat{x}[m, n] - x[m, n] \quad (2)$$

At each iteration the weights  $w_{(m,n)}[k, l]$  are adapted using a two-dimensional LMS-type adaptation algorithm:

$$w_{(m+1,n)}[k, l] = w_{(m,n)}[k, l] + \mu \times e[m, n] \times x[k, l] \quad (3)$$

where  $(k, l) \in \mathcal{R}$ , the region of support, and  $\mu$  is the adaptation constant.

Since microcalcifications are isolated bright spots, the prediction error sequence deviates from Gaussianity around microcalcification locations. Therefore, a statistics of the prediction errors is computed to determine whether they are samples from a Gaussian distribution. The regions with Gaussianity test values higher



Figure 1: Part of a mammogram image which contains a microcalcification cluster.

than a set threshold value,  $T_h$ , are marked as regions of microcalcification clusters. Figure 1 shows part of a mammogram image containing a microcalcification cluster. In Figure 2, the detection scheme output is given.

The contribution of this paper is twofold: we propose an alternative Gaussianity test to the OCM Gaussianity test. We proceed by computing the sensitivity curves for the two techniques. The sensitivity curves indicate that our test is more sensitive to outliers therefore it provides higher microcalcification detection rates. The results are validated by computer simulation studies. In the following section, the OCM Gaussianity test is reviewed. Section 4 describes the design of a new Gaussianity test based on the sensitivity curve concept reviewed in Section 3. Results and conclusions are given in Section 5.

## 2. THE OCM GAUSSIANITY TEST

The OCM Gaussianity test is based on the sample estimates of the first three moments  $I_1, I_2, I_3$  of the prediction errors. Estimates of the moments are given by:

$$I_k = \frac{1}{M \times N} \sum_{m=1}^M \sum_{n=1}^N e^k[m, n], \quad k = 1, 2, 3 \quad (4)$$

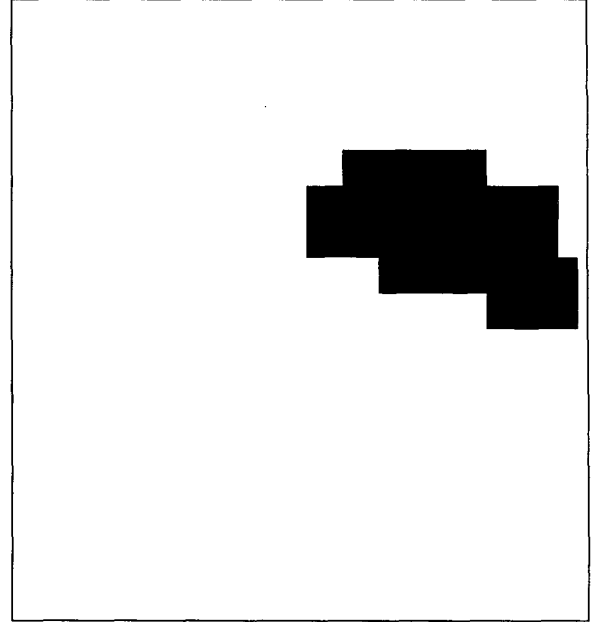


Figure 2: Detection scheme output. Regions with microcalcifications are indicated by the detection scheme.

where,  $e[m, n]$ 's ( $m = 1, \dots, M, n = 1, \dots, N$ ) are individual error values at the location  $(m, n)$  after adaptive filtering and  $M \times N$  is the total number of error pixels in the square region ( $M = N = 30$  in our experiments). For Gaussian distributed sequences,  $I_1, I_2, I_3$  converge to the following values as  $M, N$  go to infinity under the ergodicity assumption

$$I_1 \rightarrow \mu, \quad I_2 \rightarrow \sigma^2 + \mu^2, \quad I_3 \rightarrow \mu^3 + 3\sigma^2\mu \quad (5)$$

where  $\mu$  and  $\sigma^2$  denote the mean and the variance of the error sequence  $e$ , respectively. With these limit values, the nonlinear expression

$$h(I_1, I_2, I_3) = I_3 - 3I_1(I_2 - I_1^2) - I_1^3 \quad (6)$$

is close to zero for Gaussian distributed sequences. Otherwise, it is concluded that the sequence deviates from Gaussianity. In the following section we estimate the sensitivity of this Gaussianity test to outliers.

## 3. SENSITIVITY ANALYSIS

The sensitivity analysis of Gaussianity tests are based on sensitivity curve which is a finite-sample version of the influence function. The Influence Function (IF) of an estimator,  $T$ , for the cumulative distribution,  $F$ , is

given by [6]:

$$IF(x; T, F) = \lim_{t \rightarrow 0} \frac{T((1-t)F + \delta x) - T(F)}{t} \quad (7)$$

where  $\delta x$  is the probability measure which puts mass 1 at point  $x$ . The influence function describes the effect of an infinitesimal contamination at the point  $x$  on the estimate. The influence function for the sample mean function,  $T_n = \frac{1}{n} \sum_{i=1}^n x_i$ , for the Gaussian distributed sequences is  $IF(x; T, F) = x$ .

Tukey derived a simple finite-sample version of Equation 7 [7]:

$$SC_n = n[T_n(x_1, \dots, x_{n-1}, x) - T_{n-1}(x_1, \dots, x_{n-1})] \quad (8)$$

This is called the sensitivity curve which basically examines the effect of an additional term,  $x$  as an outlier on the overall estimator. First, the estimator value,  $T_{n-1}(x_1, \dots, x_{n-1})$  for  $n-1$  terms is calculated. Next, the outlier term,  $x$  is added to the sequence and the estimator is again calculated for the  $n$  terms,  $T_n(x_1, \dots, x_{n-1}, x)$ . The difference between these two estimator values exhibits the effect of the outlier on the estimator. The sensitivity curve,  $S_n(x)$ , can be plotted against values of the outlier,  $x$  to visualize the effect of different values of outliers on the the overall sensitivity of the estimator (see Figure 3).

The Gaussianity test in Equation 6 can be simplified to obtain the following expression:

$$h(I_1, I_2, I_3) = I_3 - 3I_1I_2 + 2I_1^3 \quad (9)$$

The overall sensitivity of the Gaussianity test  $h(I_1, I_2, I_3)$  can be calculated by calculating  $SC_n(x)$  values for  $I_3, I_1I_2, I_1^3$  and then combining the results. So, the overall sensitivity curve is

$$SC_n(x) = x^3 \left(1 - \frac{3}{n} + \frac{2}{n^2}\right) + x^2 \varphi \left(-\frac{3}{n} + \frac{6}{n^2}\right) + x \left(-\frac{3\Omega}{n} + \frac{6\varphi^2}{n^2}\right) \quad (10)$$

where  $\varphi := x_1 + \dots + x_{n-1}$ , and  $\Omega := x_1^2 + \dots + x_{n-1}^2$ .

For Gaussian sequences and for large values of the sample size,  $n$

$$\frac{\varphi}{n} \rightarrow \mu, \quad \frac{\Omega}{n} \rightarrow (\mu^2 + \sigma^2) \quad (11)$$

where  $\mu$  is the mean and  $\sigma$  is the standard deviation. Substituting these values into Equation 10, we get

$$SC_n(x) = x^3 \left(1 - \frac{3}{n} + \frac{2}{n^2}\right)$$

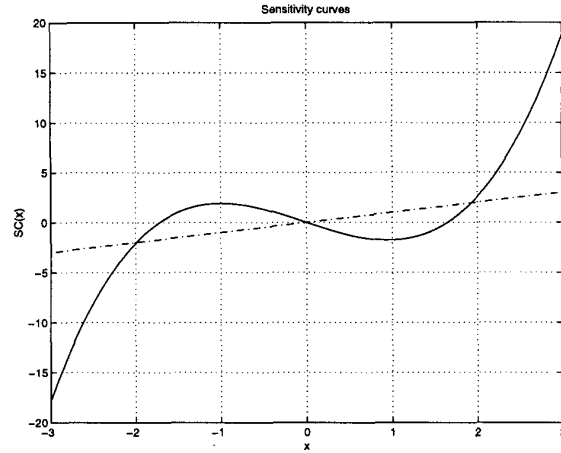


Figure 3: Sensitivity curves for the mean (dashed lines) and for the Gaussianity test in Equation 9 (solid line)

$$+ x^2 n \mu \left(-\frac{3}{n} + \frac{6}{n^2}\right) + x \left(-\frac{3n(\mu^2 + \sigma^2)}{n} + \frac{6n^2 \mu^2}{n^2}\right) \quad (12)$$

For Gaussian sequences with zero mean and standard deviation of one, Equation 12 further simplifies to:

$$SC_n(x) = x^3 \left(1 - \frac{3}{n} + \frac{2}{n^2}\right) - 3x \quad (13)$$

For large values of  $n$  (i.e.,  $n \rightarrow \infty$ ), the sensitivity curve is reduced to

$$SC_n(x) = x^3 - 3x \quad (14)$$

If we choose  $x_1, \dots, x_{n-1}$  as 900 random Gaussian distributed numbers with zero mean and standard deviation of one, then the sensitivity curve in Figure 3 is obtained. This curve closely fits to the curve  $y = x^3 - 3x$  as can be expected from Equation 12.

It is possible to design other Gaussianity tests, which makes use of higher order moments in order to have higher sensitivities. This will be useful in the detection of microcalcifications as they will be treated as outliers and more sensitive tests will be able to detect them with more ease.

#### 4. FOURTH ORDER GAUSSIANITY TEST

Traditionally, both third and fourth order statistical parameters are used in Gaussianity detection. In the OCM test, parameters up to the third order are used. By introducing the fourth order, the sensitivity of the

statistical test to outliers can be improved. Since microcalcifications will produce outliers in the error image and tests with higher sensitivities can detect the outliers better, the higher the sensitivity of the Gaussianity test, the better its microcalcification detection performance is.

The fourth order moment is derived from the moment generating function,  $M_X(t)$ , of the Gaussian distribution [8]

$$M_X(t) = e^{t\mu + \sigma^2 t^2/2} \quad (15)$$

The  $k^{th}$  order moment of distribution,  $I_k$ , is defined in terms of the moment generating function as follows

$$I_k = E(x^k) = \frac{d^k}{dt^k} M_X(t)|_{t=0} \quad (16)$$

Of particular interest here is the fourth moment which is obtained using the following relation:

$$\begin{aligned} I_4 &= [\mu^4 + 3\sigma^4 + 6\mu^2\sigma^2 + t(12\mu\sigma^4 + 4\mu^3\sigma^2) \\ &+ t^2(6\sigma^6 + 6\mu^2\sigma^4) + t^3(\mu\sigma^6 + 3\mu\sigma^6) \\ &+ t^4\sigma^8]e^{t\mu + \sigma^2 t^2/2} \end{aligned} \quad (17)$$

In the limit, when the value of  $t$  is taken as zero, only the first three terms remain in the above expression and these constitute the fourth moment,  $I_4$ . Hence,

$$I_4 = E(x^4) = \mu^4 + 6\mu^2\sigma^2 + 3\sigma^4 \quad (18)$$

In designing the test, we want to establish a function such that it will assume the value of zero for Gaussian distributed sequences. First, a term is needed to eliminate the  $\mu^4$  term. In the limit, the moment  $I_1$  approaches to the value of  $\mu$ , therefore, the fourth power of this moment can be subtracted from the fourth moment term,  $I_4$ . In the limit,  $(I_2 - I_1^2)$  approaches to  $\sigma^2$ , which can then be used to eliminate the second and third terms of the moment expression. Therefore, the statistic for the Gaussianity test turns out to be:

$$H(I_1, I_2, I_4) = I_4 - I_1^4 - 6I_1^2(I_2 - I_1^2) - 3(I_2 - I_1^2)^2 \quad (19)$$

which can be then simplified by eliminating the repetitive terms to get:

$$H(I_1, I_2, I_4) = I_4 + 2I_1^4 - 3I_2^2 \quad (20)$$

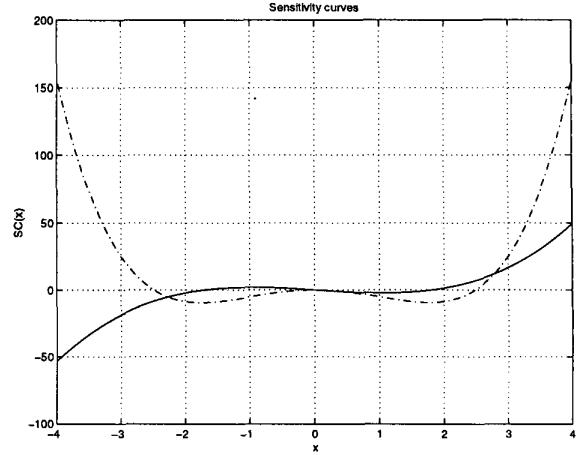


Figure 4: Sensitivity curves for the Gaussianity test in Equation 9 (solid line) and for the Gaussianity test in Equation 20 (dashed line).

The overall sensitivity curve of the new test is:

$$\begin{aligned} SC_n(x) &= x^4 \left(1 - \frac{3}{n} + \frac{2}{n^3}\right) + x^3 \varphi \left(\frac{8}{n^3}\right) \\ &+ x^2 \left(\frac{8}{n^3} \varphi^2 - \frac{6}{n} \Omega\right) \\ &+ x \left(-\frac{4\varphi^3}{n^3}\right) \end{aligned} \quad (21)$$

As  $n \rightarrow \infty$ ,  $\frac{\varphi}{n} \rightarrow \mu$ ,  $\frac{\Omega}{n} \rightarrow (\sigma^2 + \mu^2)$ , Equation 21 becomes:

$$\begin{aligned} SC_n(x) &= x^4 + x^3 \left(\frac{8}{n^3} n \mu\right) \\ &+ x^2 \left(\frac{8}{n^3} n^2 \mu^2 - \frac{6}{n} n (\sigma^2 + \mu^2)\right) \\ &+ x \left(-\frac{4}{n^3} n^3 \mu^3\right) \end{aligned} \quad (22)$$

which boils down to the following relation for Gaussian signals with zero mean and the standard deviation of one:

$$SC_n(x) = x^4 - 6x^2 \quad (23)$$

So, the overall sensitivity of the newly developed Gaussianity test is higher than that of Equation 9. As a natural extension, other Gaussianity tests which use moments higher than fourth order moments can be designed. However, these moments converge very slowly to the normal distributions and so should not be used unless very large samples are processed [5].

Test Statistic	Mean	Minimum	Maximum
$h(I_1, I_2, I_3)$	38.6	8.4	306.0
$H(I_1, I_2, I_4)$	406.8	52.4	3712.3

Table 1: Test statistics in regions with microcalcifications.

Test Statistic	Mean	Minimum	Maximum
$h(I_1, I_2, I_3)$	0.4	-2.1	2.2
$H(I_1, I_2, I_4)$	1.25	-4.4	11.1

Table 2: Test statistics in regions without microcalcifications.

## 5. RESULTS AND CONCLUSIONS

When the new Gaussianity test is used with the adaptive filtering scheme, the statistics in Tables 1 and 2 are obtained. The test results are obtained from 100 different regions on 5 different mammogram images. With this test, the effect of outliers is more apparent. Figure 4 shows this effect. Since the microcalcifications appear as outliers, they will be more pronounced with this test. Actually, the values of the newly developed test in Equation 20, reflect this change, while the values in regions with no microcalcifications remain close to zero, in regions with microcalcifications, both Gaussianity tests produce high test statistics values. The fourth order Gaussianity test gives higher values than the OCM Gaussianity test.

As the range between the maximum value of one region and the minimum value of the other region is larger, it is possible to set the detection threshold,  $T_h$ , at a higher level and eliminate some of the false alarms (or single-bright spot regions). We tested the effectiveness of our new Gaussianity test using the Nijmegen mammogram image database<sup>1</sup>. The database contains 40 digitized mammogram images. Using only the OCM test in our previous detection scheme we were able to get 1.4 false alarm regions per image when all the radiologist-approved microcalcification clusters were detected [1]. When the newly developed Gaussianity test is used in combination with the OCM Gaussianity test, the false alarm rate decreases from 1.4 per image to 1.125 per image.

<sup>1</sup>Images were provided by courtesy of the National Expert and Training Centre for Breast Cancer Screening and the Department of Radiology at the University of Nijmegen, the Netherlands.

## 6. REFERENCES

- [1] M. Nafi Gurcan, Yasemin Yardimci, A. Enis Cetin, "Microcalcification Detection Using Adaptive Filtering and Gaussianity Tests," Proceedings of the Fourth International Workshop on Digital Mammography, pp. 157-164, June 1998, Nijmegen, The Netherlands.
- [2] Metin Nafi Gurcan, Y. Yardimci, A. E. Cetin, "2-D Adaptive Filtering Based Gaussianity Tests in Microcalcification Detection," Proceedings of SPIE Visual Communications and Image Processing Conference, vol. 3309, part II, pp. 625- 633, 24-30 January, 1998, San Jose, CA.
- [3] M. Nafi Gürcan, Yasemin Yardımcı, A. Enis Çetin, Rashid Ansari, "Detection of Microcalcifications in Mammograms Using Higher Order Statistics," IEEE Signal Processing Letters, vol. 4, no. 8, pp 213-216, August 1997.
- [4] R. Ojeda, J. Cardoso, E. Moulines, "Asymptotically Invariant Gaussianity Test For Causal Invertible Time Series," Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 5, pp. 3713-3716, April 21-24, 1997.
- [5] E. Moulines, K. Choukri, "Time-Domain Procedures for Testing that a Stationary Time-Series is Gaussian," IEEE Transactions on Signal Processing, vol. 44, no. 8, pp. 2010-2025, August 1996.
- [6] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, W. A. Stahel, Robust Statistics: The Approach Based on the Influence Functions, John Wiley & Sons, New York : 1986.
- [7] J. W. Tukey, Exploratory Data Analysis, Addison-Wesley, Reading, Mass. : 1971.
- [8] W. W. Hines and D. C. Montgomery, Probability and Statistics in Engineering and Management Science, John Wiley and Sons, New York : 1980.