

Detection of Microcalcifications in Mammographies Based on Linear Pixel Prediction and Support-Vector Machines

F. Martínez-Álvarez
Univ. Sevilla
fmartinez@lsi.us.es

A. Troncoso
Univ. Pablo Olavide
ali@upo.es

J. C. Riquelme
Univ. Sevilla
riquelme@lsi.us.es

J. S. Aguilar-Ruiz
Univ. Pablo Olavide
aguilar@upo.es

Abstract

Breast cancer is one of the diseases causing the largest number of deaths among women. Its early detection has been proved to be the most effective way to combat it. This work is focused on developing an integral tool able to detect microcalcifications in mammographies, since the presence of these particles is a clear symptom of an incipient cancer. The proposed approach combines two techniques successfully used in other areas separately, such as linear pixel prediction and support-vector machines, in order to obtain almost perfect prediction accuracy. Moreover, a filter has been designed with the aim of decrease the processing time. The result verges on 96% of hits, improving previous works by 6%, on average.

1. Introduction

Image processing techniques have been developed during the last two decades so to help in the diagnosis of the breast cancer, since the early detection of the cancer increases the survival rate from 58% to 86% five years after its appearance [1]. Thus, monitoring breast cancer programs have turned into an essential tool the use of which should be compulsory for most professionals.

The calcifications are usually present in mammographies. Some of them are the result of a cellular secretion whereas others may be the consequence of the necrotic cellular detritus. In the same way, they may appear as a response to an inflammation, traumatisms or radiation. Typical calcifications associated with cancer are small, with a standard size of about 0.5 millimetres [2]. The calcifications found in mammographies have a size from 0.2 to 0.3 millimetres and are called *microcalcifications*.

The main objective of a breast cancer detection system is finding microcalcifications. Researchers are usually worried about optimizing only one part of the general outline of prediction and, in that sense, it is desirable to design a tool which chose the optimum algorithm for each part and integrate them.

Few years ago, numerous studies centred on restrictive systems with the intention of isolating the microcalcifications from the rest of the mammography [3]. Nevertheless, recent approaches based on neural networks [4] or wavelet transforms [5], have obtained regions of decision appreciably more complex.

To achieve this aim, it is necessary to perform a previous task: extracting and selecting features in the neighbourhood of the possible microcalcification [6]. It is in densest mammographies where this task becomes particularly difficult because the pixels surrounding the microcalcifications have a similar intensity value. The filter designed is especially effective to ease this effect.

The rest of this paper is divided as follows. In Section 2, the new filter is presented. It also provides an explanation of the linear prediction algorithm and the support-vector machine. Section 3 includes the general procedure and shows the improvement achieved by this proposal. Section 4 shows the final aspect of a picture after its processing. Finally, Section 5 exposes the conclusions and gives some advices on how this approach should evolve.

2. Improving the algorithm

2.1 Use of the filter

The starting point is the algorithm proposed in [7] and improved in [8]. It assumes that the density and texture of the breast is uniform. Only the zones that do not fulfil both requirements will be considered pre-candidates pixels. Thus, the algorithm predicts the grey value of a certain pixel, given the value of the pixels round it. If the difference between the real and predicted values is high, the pixel is marked as a pre-candidate.

This technique gave surprisingly good results, as no filtering was done before the prediction. However, the processing time was extremely high. Considering that only a third of the image provides useful information, a filter *ad hoc* has been proposed in order to extract the information belonging to the breast from the image and to remove the offset existing in the image. The code of the filter is shown in Figure 1.

```
if (image[i][j] <= BACKGROUND){
    for(i=0;i<rows;i++){
        for(j=0;j<columns;j++){
            if (image[i][j] <= THRESHOLD) {
                aux += image[i][j];
                number_pixels++;
            }
        }
        percent = 100*number_pixels/(columns*rows);
        if (percent > SHADE)
            average = floor(aux/number_pixels) + DARK;
        else
            average = floor(aux/number_pixels) + CLEAR;
    }
}
```

Figure 1. Source code of the filter implemented.

If a pixel has a value lower than BACKGROUND it is considered to be useful information of the image. Otherwise, it is considered to belong to the background of the picture and avoided in further steps.

As for the THRESHOLD, it is to decide whether the image is dark or not. Pixels under this threshold are considered to belong to regions abnormally clear.

The variable percent represents the percentage of pixels that have a value lower than the THRESHOLD. Those mammographies with a value higher than SHADE in their percent variable were considered dark and, otherwise, clear. This fact is of utmost importance because the average variable is used to delete the offset existing in the image in the rest of the process. Both DARK and CLEAR are parameters whose election is justified because it helps to distinguish between a dark and a clear image and the response of the algorithm will be different in these two cases.

The database analyzed has pixels encoded by twelve bits, this is to say, 4096 levels of grey. All BACKGROUND, THRESHOLD, SHADE, DARK and CLEAR values have been selected with the help of an oncologist and their values are 3500, 2000, 0,17, 1300 and 700.

2.2. Selecting microcalcifications

Once the pre-candidates pixels are selected, the Tail-Ratio parameter [6] is used as a grade of reliability of the selected pre-candidates pixels because it tests the histogram of the pixels. Since the microcalcifications are points with a high intensity, the probability density function

(PDF) of the pixels round a microcalcification has a right tail longer than the left one. The direct application of this parameter performed a satisfying sift: the number of candidates is the 12% of the number of pre-candidates.

The final step entails deciding if the candidates pixels are centroids of microcalcifications. The method selected is a support-vector machine classifier [9].

The input parameters for the classifier are some features selected from an initial set. Table 1 shows the features which make up this set. The formula of these features can also be found in [6].

Table 1. Initial set of features

Abbreviation	Description
TR	Tail-Ratio: grade of reliability of the pre-candidates
K	Size of the surrounding square whose pixels exceed the TR parameter
ID	Inter-Distance: pixels with intensity higher 98% will belong to microcalcification
AH	Average Height of the histogram in the neighbourhood $k \times k$
E	Entropy
C	Contrast
DR	Dynamic Range: difference of grey in the square $k \times k$
CG	Correlation Gaussian distribution: assumed to be circular and symmetric

The fact of using many descriptors does not imply a better classification. Contrary to the expectations, it only increases the complexity of the system and for this reason it is essential to obtain an optimal set of features for subsequent steps. Two methods, proposed in [10], were followed in order to obtain this set: sequential backward selection (SBS) and sequential forward selection (SFS). The result of applying these two methods to the initial set is shown in Table 2.

Table 2. Application of SFS and SBS to the initial set of features

Method	Feature set	Average error
SFS	TR, C, DR, ID	4.27%
SBS	AH, ID, MS, TR	5.31%

Thus, the four features provided by the SFS method have been chosen because it has an average error lower than the one given by the SBS method.

3. Procedure

Figure 2 shows a detailed procedural schema of the full system. Thus, the input file is transformed into a *bmp* file with the microcalcifications well marked. A *log* file is also generated and its benefit is to save all the possible problems occurred during the execution of the program and it also shows all the meaningful information of the picture. The *svm_parameters.m* file contains a previous knowledge of the features selected by the SFS method.

3.1 Parameters of quality

Neither all the existing microcalcifications are detected, nor all the detected ones are microcalcifications. Thus, a microcalcification can be classified into four possible cases. The *true positives* (TP) represent the number of microcalcifications properly detected at the end of

the process. However, the *false positives* (FP) is related to those pixels which have been considered to be microcalcifications by the tool but they are not. As for the *false negatives* (FN), it indicates the number of microcalcifications which have been ignored by the algorithm. Finally, the *true negatives* (TN) are the number of candidates properly discarded by the support-vector machine.

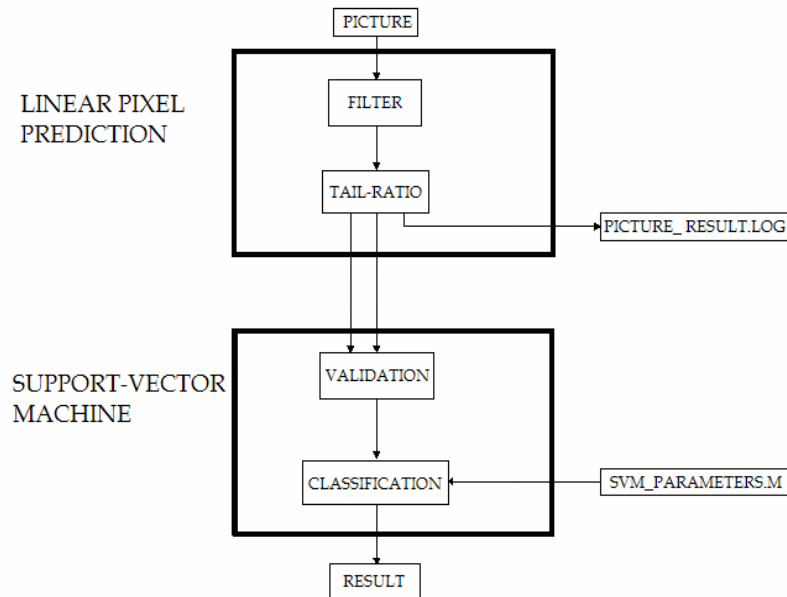


Figure 2. Diagram of the system. Two phases clearly differentiated: linear pixel prediction with filtering and classification by means of the support-vector machine.

3.2 Parameters of measurement

In order to have a precise evaluation of the results, some parameters are defined. The *sensitivity* is the probability to detect a microcalcification properly. Its formula is:

$$Sensitivity = \frac{TP}{TP + FN} \quad (1)$$

Other parameter is the *specificity* which is the ratio of candidates properly discarded by the support-vector machine.

$$Specificity = \frac{TN}{TN + FP} \quad (2)$$

The *positive predictive value* (PPV) is the probability that a detected microcalcification is a real one. Its formula is:

$$PPV = \frac{TP}{TP + FP} \quad (3)$$

The *negative predictive value* (NPV) is the probability that a discarded microcalcification is not a real one. Its formula is:

$$NPV = \frac{TN}{TN + FN} \quad (4)$$

4. Results

The proposed algorithm has been successfully applied to 94 mammographies obtained from a database supplied by a private hospital.

The initial results were quite good, with a sensitivity of nearly 90% [8]. Nevertheless, after optimizing the source code by adding a filter, the results have been improved and the processing time has been deeply decreased. It is especially remarkable that only two mammographies have been improperly detected with a sensitivity of 88%. Both of them share the common feature: they have an extremely high grey level. To be precise, only the 15% of the pixels belonging to these mammographies have a grey average level below the threshold defined, whereas the other mammographies examined have a 24% of the pixels over the threshold, on average. Therefore, the algorithm does not respond properly to abnormally dark mammographies.

The algorithm does not detect big microcalcifications. There are two main causes that justify this voluntary limitation. From one side, a big microcalcification is clearly visible for specialists and its automatic detection is useless. On the other hand, detecting them would involve a considerable increase of the processing time.

Table 3 shows the results obtained from the application of the algorithm to the mammographies of the database analysed. All of them have been enhanced. The feature that has been notably improved is the PPV: from an initial 89% of accuracy to a 97%. Those microcalcifications whose perception with just a look was not possible are now detected, thanks to the extraordinary sensitivity achieved. Therefore, better detections are done in a third of time, as it can be seen in Table 3.

Table 3. Comparison between using or not the filter.

Parameters	Without filter	With filter
Sensitivity	90.21%	95.87%
Specificity	97.67%	99.37%
PPV	89.15%	97.08%
NPV	96.27%	99.31%
TP (average)	93.1	106.3
FP (average)	13	2.3
FN (average)	6.7	4.9
TN (average)	1038.6	821.1
Candidates (average)	1151	934
Processing time (average)	83"	28"

Figures 3.a – 3.c show the changes suffered by the pictures during the process of detecting. Note that only the parts of interest of the mammography are shown. Figure 3.a represents the mammography as it was before the detection. Figure 3.b shows the pre-candidates (big circles in black) marked on the mammography after applying the filter. It also shows the candidates calculated by the TR parameter (inside a white frame). Figure 3.c shows the microcalcifications detected by the support-vector machine.

5 Conclusions

The goal of the paper was to develop an integral tool able to detect microcalcifications in mammographies accurately. This objective has been widely achieved insofar as the integration of two different techniques, filtered linear pixel prediction and SVM classifier, has allowed a sensitivity of nearly 96% in their detection.

The false positives have traditionally been the factor to be decreased since sensitivity and specificity had already acceptable values. This work has also reached this objective: from 13 to 2.3 per image, on average.

Although neural networks have had good results, it has been proved that the use of others well-known techniques properly tuned may be useful in image processing.

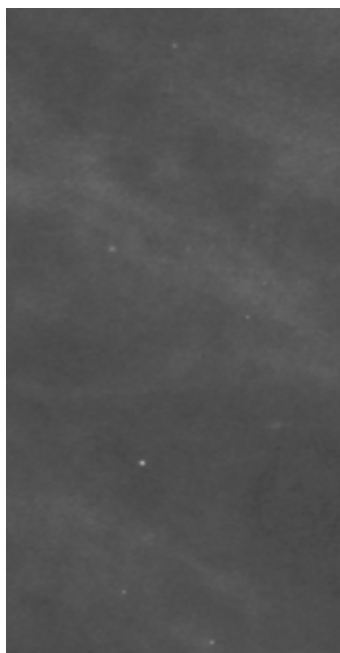


Figure 3.a. Mammography

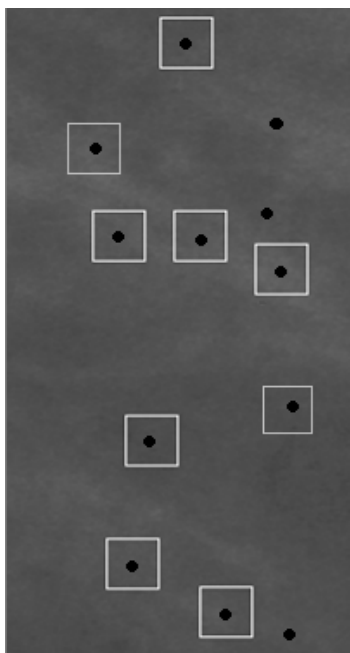


Figure 3.b. Candidates

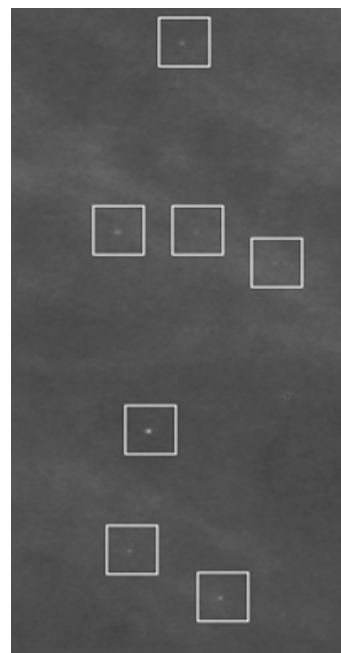


Figure 3.c. Detection

Acknowledgements. The authors would like to acknowledge the financial support from the Spanish Ministry of Science and Technology, project TIN2004-00159 and from the Junta de Andalucía, project P05-TIC-00531. The help provided by Dr. Mohedano, oncologist of the *Ciudad de Jaén* Hospital, in the interpretation of the mammographies and the results has been crucial in the develop of this research.

References

- [1] Zhou, X., Gordon, R.: Detection of early breast cancer: an overview and future prospects. *Critical Reviews in Biomedical Engineering*, Vol. 17. (1999) 203- 255
- [2] Cowen, A. R., Launders, J. H., Jadav, M., Brettle, D. S.: Visibility of microcalcifications in computed and screen-film mammography. *Physics in Medicine and Biology*, Vol. 42. (1997) 1533–1548
- [3] Gavrielides, M. A., Lo, J. Y., Floyd, C. E.: Parameter optimization of a computer-aided diagnosis scheme for the segmentation of microcalcification clusters in mammograms. *Medical Physics*, Vol. 29, Num. 4, (2002) 475-483.
- [4] Papadopoulos, A., Fotiadis, D.I., Likas, A.: An automatic microcalcification detection system based on a hybrid neural network classifier. *Artificial Intelligence in Medicine*, Vol. 25, (2002) 149-167
- [5] Strickland, R. N., Hahn, H. I.: Wavelet Transforms for Detecting Microcalcifications in mammograms. *IEEE Trans. On Medical Imaging*. Vol. 15, Num. 2, (1996) 215-228
- [6] Acha, B., Serrano, M. C., Rangayyan, R. M., Leo Desautels, J. E.: Detection of Microcalcifications in Mammograms. *Recent Advances in Breast Imaging, Mammography and Computer-Aided Diagnosis of Breast Cancer*. SPIE - the International Society for Optical Engineering (2006) 297-320
- [7] Serrano, M. C., Díaz-Trujillo, J, Acha, B., Rangayyan, R. M.: Use of 2d Linear Prediction Error to Detect Microcalcifications in Mammograms. II Congreso Latinoamericano de Ingeniería Biomédica. Congreso Latinoamericano de Ingeniería Biomédica. Num. 2, (2001)
- [8] Acha, B., Serrano, M. C., Rangayyan, R. M.: Detection of Microcalcifications in Mammograms Using 2d Prediction Filtering and a New Statistical Measure of the Right Tail Weight. *Proceedings of Embecc 2005*. Ifmbe Proceedings Series. (2005) 3112-3117
- [9] Cortes, C., Vapnik, V. N.: Support-vector networks. *Machine Learning Journal*. (1995) 273-297
- [10] Reeves, S. J., Zhao, Z.: Sequential algorithms for observation selection. *IEEE Trans. Signal Process*. Vol. 1, Num. 47, (1999) 123–132