Applied Mathematical Sciences, Vol. 6, 2012, no. 131, 6533 - 6544

Detection of Microcalcification in Digitized Mammograms Using Weighted Local Differences and Local Contrast

Younes Kabbadj

LIMIARF, Faculty of sciences, Rabat, Mohammed the fifth University, Agdal, Morocco
kabbadjyounes@yahoo.fr

Fakhita Regragui

LIMIARF, Faculty of sciences, Rabat, Mohammed the fifth University, Agdal, Morocco
regragui@fsr.ac.ma

Mohammed Majid Himmi

LIMIARF, Faculty of sciences, Rabat, Mohammed the fifth University, Agdal, Morocco
himmi@fsr.ac.ma

Abstract

Breast cancer remains one of the deadliest diseases among women. Microcalcifications can be an early indicator of a breast cancer. Thus when they are present their detection during a screening test is very crucial. But due to their small size and low contrast in mammographies their detection is difficult. Therefore many computer aided diagnosis mathods have been developped to help and assist rediologist during their screening tests. This paper presents a novel approach to detect microcalcifications on digitized mammaographies using fuzzy logic and support vector machines with 9 statistical and geometric features. Our method was tested on 120 cases from the Mias database including 22 microcalcification cases, 60 normal cases and 70 abnormal cases.We have reached an **Az** Value of 0,91 making this method very comparable to the State Of Art works.

Keywords: Computer Aided Detection, Microcalcification Detection, Fuzzy Logic, Support Vector Machines

1 Introduction

Breast cancer is a widespread form of cancer around the world. This disease constitutes 22% of all cancer cases and is still one of the deadliest forms especially when it is lately detected. Despite these alarming facts these disease is very common in arabic countries and Mastectomy is still performed in more than 80% of the cases [3]which can be very traumatic, both physically and psychologically. An early detection is the key solution not only to avoid Mastectomy and improve women's survival chances but also to reduce the cost of medical care. Microcalcifications(Mcs) are small spots of calcium deposits that generally appear in clusters. The presence of Mcs can be an early sign of breast cancer, therefore their detection on mammograms is very important for radiologists. But due to their small size and the low contrast in x ray images, their detection is often missed wich can lead to negative effects. Computer aided detection(CADe) are a class of methods and procedures that help doctors in their interpretation of medical images. CADe have been used for breast masses and mcs detection since decades benefiting from improvements in the fields of applied mathematics, signal processing, image processing and pattern recognition. Many researchers have tried over the past two decades to give useful contribution in CADe for mcs detection. CADe provides any radiologist a second reading which is makes his performance higher than when he relies only on his own reading. Many efforts has been made to create mcs detection systems. They generally consist of three blocks :preprocessing, regions of interest specification and finally detection. In the work of Wang et al. [18], they introduced wavelet for mcs detection by decomposing images into frequencies subbands and reconstructed only from high frequencies. To isolate mcs they segmented the reconstructed images. Sakka et al. used high frequencies from symlet,haar and deaubechies wavelets to isolate mcs components and then reconstructed the image from those components [12]. They evaluated their method using horizontal, vertical and diagonal details energies. Soltanian et al. compared multiwavelets, wavelets, Haralick, and shape features for mcs classification in mammograms using genetical algorithms and Roc curves [14]. The best results were found with multi wavelet features. Papadopoulos et al. tested image processing and wavelet techniques performance in mcs enhancement [10]. The target performance is related the percentage of the most contrasted pixels and the size of the minimum detectable object which could satisfactorily represent a mcs. The best results were obtained using local range modification and wavelet linear stretching. Same authors in [11] used breast region localisation,logical addition of two binary images resulting from a high frequency response filter and the other from background substraction to provide suspicious mcs. Resulting Objects from this phase which are smaller than a predetermined size are eliminated to reduce false positive. A classification module based on principal component analysis for feature selection associated with an artificial neural network is finally used and its output are the detected mcs. Eddaoudi et

al. [2] used a texture coding for mcs detection.Instead of processing the original, a locally coded image with a limited number of gray levels is found using a local histogram, then the least represented gray levels are replaced by the most represented gray levels. The procedure reduces computational cost. Haralick features from gray level co-occurrence matrix are computed to feed an Svm classifier to finally detect mcs from the processed image.
In this paper, we propose a method for microcalfication detection based on optimal thresholding, fuzzy logic and a support vector machine. The following section describes a algorithm for breast region isolation. Section 3 presents the fuzzy inference system then Section 4 is devoted to the Support Vector Machine followed by Section 5 which presents and discusses the results. The last section is a general conclusion.

2 Breast region isolation :

Breast images do not only include breast region. A digitized x-ray mammography consists of a rectangular region that includes the whole breast surface breast artifacts. The rest of the image is not covered by the breast. Since when performing screening and computer aided diagnosis, non-breast regions are meaningless, it seems natural and computationally efficient to eliminate those regions. In order to identificate the breast region, we perform this with a three phase algorithm :
-To separate bright parts from dark ones on the screen film,we perform a global thresholding on the whole image using the well known and efficient Otu's Thresholding [9]. Otsu's thresholding is an efficient, simple and widely used method when dealing with images of objects and homegeneous background which is the case for our study. It separates the image pixels into two groups depending on the optimal threshold which is deduced from histogram statistics, this method searches over all possible thresholds which is the one that minimizes the intra-classes variance defined in equation(1)

$$\sigma_p^2(t) = P_1(t)\sigma_1^2(t) + P_2(t)\sigma_2^2(t) \tag{1}$$

where $P_i(t)$ are the probabilities of the classes and $\sigma_i^2(t)$ their respective variances.

-Given the output of Otsu's thresholding which was originally developed to identify pixels belonging to background and those belonging to the object of interest, it is crucial to perform an further processing in order to identify breast from screen film artifacts and other objects that do not belong to the breast. In order to identify the breast it seems suitable to use a connected component labelling algorithm to attribute to each a pixel a label which determines the region to which it belongs. According to authors in [16] there are among others two main approaches to perform Connected Component Labelling : a category that repeats passes through the image attributing labels in backward and forward until no label changes [6, 7]

and an other category that performs one pass to assign provisional labels and an other pass to resolve the label equivalences [4, 5]. Since the second category is clearly the computationally most optimal one, we used an algorithm from this category. -Given that breast region represents the biggest part on the image and we have already grouped pixels from the same regions using Connected Component labelling, we have only to count for each label(each region), the number of pixels that belong to and then identify the region with the biggest number of pixels which is the breast region.

3 Fuzzy inference system

Radiologists often rely on mcs appearance to locate them on mammograms. Mcs contrast and luminosity in the images distinguish them from normal breast tissue. However these two characteristics have different meanings depending on the person who is performing the screening. Furthermore the poor contrast in mammograms especially when the breast is dense and when mcs have a very small size. In our attempt to characterize mcs in a reliable manner we have used two measures: **Weighted local differences** and **Local contrast**.

3.1 Weighted local differences

Since mcs pixels show a slightly bigger luminance than that of their surrounding neighbours we have developed an original measure that describes for each pixel in the image its level of luminance compared to its neighbours. The measure automatically gives more weight to rare combinations of gray levels which often correspond to mcs. Figure(1) illustrates High and Low membership functions.

$$ld(i,j) = \sum_{i'=-2}^{2} \sum_{i=-2}^{2} w((i,j);(i',j')) * (I(i,j) - I(i',j')) \tag{2}$$

where

$$w((i,j);(i',j')) = (1/P_{occ}((I(i,j),I(i',j'))/Norm$$

and

$$Norm = \sum_{i=-2}^{2} \sum_{j=-2}^{2} w((i,j);(i',j'))$$

$P_{occ}(I(i,j),I(i',j'))$ is the probability of co-occurence of the gray levels $I(i,j)$ and $I(i',j')$

3.2 Local contrast

Beside *Weighted local differences* we have used another measure to characterize mcs inspired by another measure originally used to evaluate the breast density [13]

$$c(i,j) = \frac{1}{25}\sqrt{\frac{\sum_{i=-2}^{2}\sum_{j=-2}^{2}(I(i,j)-I(i',j'))^2}{I(i,j)}} \tag{3}$$

3.3 Fusion of Local Contrast and weighted local means in a fuzzy inference system

In order to fully take advantage from both of the previous measures, we have built a fuzzy inference system [1] using **If**,**And** and **then** statements.
-First each function is fuzzyfied and transformed from a measure function to a membership function. Then for each input image fuzzyfied is obtained using the following inference system :

IF **ld(i,j)** is High **AND** cst(i,j) is high **THEN** mc(i,j) is high IF **ld(i,j)** is Low **OR** cst(i,j) is low **THEN** mc(i,j) is low

ld(i,j),cst(i,j) are the values for local contrast and weighted local differences and *mc(i,j)* is the membership function to the output class (microcalcification).

The membership functions high and low are mapped from any real value using the following equations:

$$High(x) = \begin{cases} 0 \;\; if \;\; x < a, \\ \frac{x-a}{b-a} \;\; if \;\; a \leq x < b, \\ 1 \;\; if \;\; x \geq b. \end{cases} \tag{4}$$

$$Low(x) = \begin{cases} 1 \;\; if \;\; x < a, \\ 1 - \frac{x-a}{b-a} \;\; if \;\; a \leq x < b, \\ 0 \;\; if \;\; x \geq b. \end{cases} \tag{5}$$

Parameters a and b were chosen as $a = 0.5Vmax$ and $b = 0.75Vmax$ with$Vmax$ the maximum value of the measure.

4 Mcs detection using Support Vector Machines

4.1 Learning Database extraction

Prior to learning one must extract relevant and meaningful features in order to build a classifier wich is our case SVMs. The classifier would serve in future simulations on new cases to classify them to 2 or more classes on the basis of values that each

sample has for each feature. In order to have a strong learning database that maximises the classification rate while minimising the number of false positives one must be very careful when choosing learning samples that would serve as entries in the learning database but also when choosing the features themselves. Some approaches chose features extracted from wavelet transforms while others extract feature directly from the original images or their enhanced versions. In the case where features are extracted from transforms such as wavelets no regions are located in the image, in this case features are extracted from the whole transform. Furthermore feature extraction can be either done for each pixel or for a set of connected pixels belonging to the same object. While single pixel feature extraction seems meaningless when dealing with microcalcifications which generally consist of several pixels and that it is weaker when pixels are noisy, pixels set feature extraction seems more suitable for our case. Thus we extracted features from manually segmented regions corresponding to mcs and negative regions with similar appearance. To fully handle characteristics of mcs we have used both geometric and statistical features. The first category provides information about the shape while the second one computes statistics based on mcs pixels that discriminate positive from negative regions.

4.2 Svm classifiers

Support Vector machines are a class of supervised classifiers developed By Vapnik [17]in the 1990s.Their strength lies on their ability to learn form high dimensional data and on their theoretical and practical strength. In the case of linearly separable classes the goal is to find the separating hyper-plan with the biggest margin.Otherwise a transformed-feature where our data is supposed to be linearly separable is generated using a kernel function. The output of the fuzzy-inference system is a limited set of regions which are suspected of being microcalcifications.This set actually may contain true mcs(true positives)beside false mcs(false positives). We used an Svm with three different kernel functions. We have trained our learners on a database consisting of 143 mcs and 166 normal regions.

For the learning and classification phases we have used 5 statistical and 4 geometrical features as described. Geometrical features include :**Eccentricity , Major Axis Length, Minor Axis Length.** Statistical features include : **Mean Gray level,Variance, Skewness, Kurtosis and Range=Max-Min**.

5 Results

To test our approach on experimental cases we have chosen the well known MIAS Database [15]. MIAS DataBase consists of 322 mammograms including normal and abnormal cases. Abnormal cases include mcs, masses , architectural distortions and asymmetries. All the cases can be divided into three categories representing

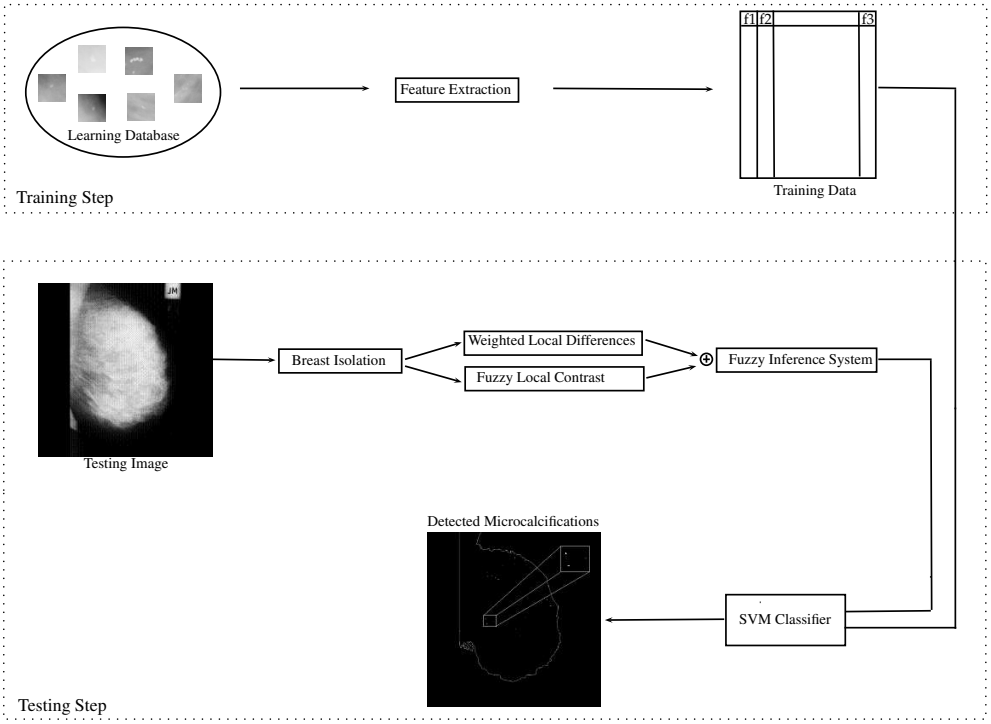


Figure 1: A Flowchart summarizing the proposed method for Mcs Detection

the density of breast tissue which are : fatty, fatty glandular and dense glandular making this database a very good sample for the validation of CADe validation. To experiment our approach we have used 120 cases with 22 mcs cases and the rest containing 70 normal and 28 abnormal cases with other types of lesions.

MIAS DataBase contains a description file that shows the location of an abnormality and approximate radius. Twelve folder cross validation was used to evaluate the performance of this method. Hence, each mammogram appears either in the test group or in the training group. A quantitative evaluation was performed using Receiver Operating Characteristic (ROC) and free Operating characteristic(FROC) fig2. The first one plots the true positive rate in function of the false positive rate and the area under this curve represents the probability that given a positive case and a negative one the classifier output will be higher for the positive case and it is not dependant on the choice of the threshold while the second one is an analysis that is a plot of operating points showing the tradeoff between the TP rates versus the average number of false positives per image. The performance of a CADe system is generally related to the sensitivity which is the ratio between the number of true positive cases TP(positive successfully detected)and the number of positive cases in the test set while specificity is the number of true negative cases(negative cases

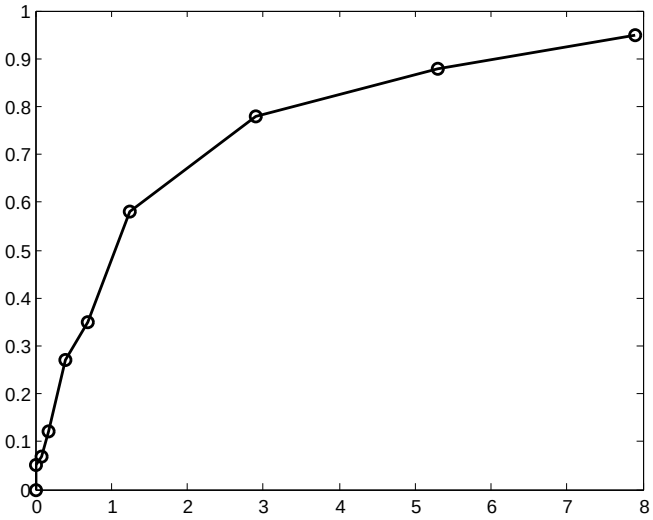


Figure 2: The resulting FROC curve for the proposed method

labelled as negative) and the total number of negative cases in the test set.

$$Sens = \frac{TP}{TP+FN} \tag{6}$$

$$Spec = \frac{TN}{TN+FP} \tag{7}$$

Generally speaking, each increase in the sensitivity decreasingly influences the specificity and vice-versa. They are both influenced by the decision threshold which can be considered as experimental parameter chosen by the user or as the influence of the random choice of the training set. The mammograms on which this analysis was performed are 1024x 1024, with a resolution of 200 microns-pixels.We have reached a sensitivity of 82,5% with an average number of false positives at 2,1 FP/image.In, the performance of our method is compared to the state of art work of Papdopoulos et al. [11] and also to the more recent work of oliver et al. [8]. Table 5 illustrates the results. The obtained value of Az is very close to its value in the other works. The FROC curve is also shown in figure. Its worth acknowledging that no post processing was used such as mcs grouping or elimination of small regions which can certainly increase the specificity. Post processing was used by Papadopoulos et al. [11] and can dramatically improve the results of any approach but is very sensitive to the choice of some parameters.

6 Conclusion

We have presented in this work a novel approach for mcs detection in digitized mammograms using breast isolation, crisp domain to fuzzy domain transformation

Work	***Az Value***	***Number of cases***
Papadopoulos et al. [11]	0,92	116
Oliver et al. [8]	0,88	112
Our proposed approach	**0,91**	**120**

Table 1: Comparison with the state of Art Approaches

and mcs localisation using a support vector machine. Breast isolation reduces the computation cost of this algorithm making it faster while fuzzy domain transformation uses two novel fuzzy measures that both handle the specific appearance of mcs on mammograms.The FROC and ROC analysis were both used as quantitative evaluation of the method. We have obtained an Az value of 0.91 making this method very similar in term of results to the state of art works without using any post-processing operations. The robustness of this method relies principally on the use of the novel fuzzy measures as well as the limited use of parameters. Future works would investigate the influence of the choice of the features on the performance of the classification module and also expand this method to the detection and classification of the other types of breast lesions.

References

[1] E. Cox, M. O'Hagan, R. Taber, and M. O'Hagen. *The Fuzzy Systems Handkbook with Cdrom*. Academic Press, Inc., 1998.

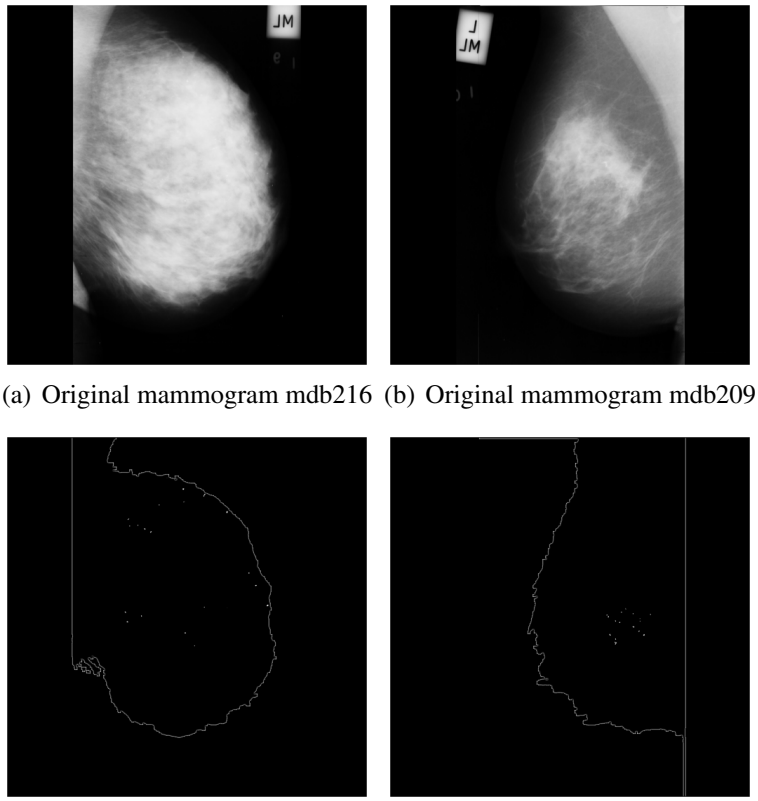
[2] F. Eddaoudi and F. Regragui. Microcalcifications detection in mammographic images using texture coding. *Applied Mathematical Sciences*, 5(8):381–393, 2011.

[3] N. El Saghir, M. Khalil, T. Eid, A. El Kinge, M. Charafeddine, F. Geara, M. Seoud, and A. Shamseddine. Trends in epidemiology and management of breast cancer in developing arab countries: a literature and registry analysis. *International Journal of Surgery*, 5(4):225–233, 2007.

[4] T. Gotoh, Y. Ohta, M. Yoshida, and Y. Shirai. Component labeling algorithm for video rate processing. In *Proc. SPIE*, volume 804, pages 217–224, 1987.

[5] T. Gotoh, Y. Ohta, M. Yoshida, and Y. Shirai. High-speed algorithm for component labeling. *Systems and Computers in Japan*, 21(5):74–84, 1990.

[6] R. Haralick. Some neighborhood operations. *Real Time/Parallel Computing Image Analysis*, pages 11–35, 1981.



(a) Original mammogram mdb216 (b) Original mammogram mdb209

(c) Results of the detection on mdb216 (d) Results of the detection on mdb209

Figure 3: This figure shows the results of the detection on two mammograms, images (a) and (b) are the original mammograms and (c) and (d) the processed images. The little white spots on (c),(d) correspond to the detected microcalcifications.

[7] A. Hashizume, R. Suzuki, H. Yokouchi, H. Horiuchi, and S. Yamamato. An algorithm of automated rbc classification and its evaluation. *Bio Medical Engineering*, 28(1):25–32, 1990.

[8] A. Oliver, A. Torrent, M. Tortajada, X. Lladó, M. Peracaula, L. Tortajada, M. Sentís, and J. Freixenet. A boosting based approach for automatic microcalcification detection. *Digital Mammography*, pages 251–258, 2010.

[9] N. Otsu. A threshold selection method from gray-level histograms. *Automatica*, 11:285–296, 1975.

[10] A. Papadopoulos, D. Fotiadis, and L. Costaridou. Improvement of microcalcification cluster detection in mammography utilizing image enhancement techniques. *Computers in biology and medicine*, 38(10):1045–1055, 2008.

[11] A. Papadopoulos, D. Fotiadis, and A. Likas. An automatic microcalcification detection system based on a hybrid neural network classifier. *Artificial intelligence in Medicine*, 25(2):149–167, 2002.

[12] E. Sakka, A. Prentza, I. Lamprinos, and D. Koutsouris. Microcalcification detection using multiresolution analysis based on wavelet transform. In *Proceedings of the International Special Topic Conference Information Technology in Biomedicine (ITAB'06)*. Citeseer, 2006.

[13] R. Sivaramakrishna, N. Obuchowski, W. Chilcote, and K. Powell. Automatic segmentation of mammographic density. *Academic Radiology*, 8(3):250–256, 2001.

[14] H. Soltanian-Zadeh, F. Rafiee-Rad, D. Pourabdollah-Nejad, et al. Comparison of multiwavelet, wavelet, haralick, and shape features for microcalcification classification in mammograms. *Pattern Recognition*, 37(10):1973–1986, 2004.

[15] J. Suckling, J. Parker, D. Dance, S. Astley, I. Hutt, C. Boggis, I. Ricketts, E. Stamatakis, N. Cerneaz, S. Kok, et al. The mammographic image analysis society digital mammogram database. 1994.

[16] K. Suzuki, I. Horiba, and N. Sugie. Linear-time connected-component labeling based on sequential local operations. *Computer Vision and Image Understanding*, 89(1):1–23, 2003.

[17] V. Vapnik. *The nature of statistical learning theory*. Springer Verlag, 2000.

[18] T. Wang and N. Karayiannis. Detection of microcalcifications in digital mammograms using wavelets. *Medical Imaging, IEEE Transactions on*, 17(4): 498–509, 1998.

Received: August, 2012