Automatic Classification of Clustered Microcalcifications by a Multiple Expert System

S. De Vito°, F. Tortorella*, M. Vento°

(°) Dipartimento di Informatica e Sistemistica
Università degli Studi di Napoli "Federico II"
via Claudio, 21 80125 Napoli – Italy
E-mail: vento@unina.it

(*) Dipartimento di Automazione, Elettromagnetismo, Ingegneria dell'Informazione e Matematica Industriale
Università degli Studi di Cassino
via G. di Biasio, 43 03043 Cassino - Italy
E-mail: tortorella@ing.unicas.it

Abstract

Mammography is a not invasive diagnostic technique largely used for early cancer detection in women breast. One of the main indicants of such disease is the presence of microcalcifications, appearing as small bright spots in the mammographic image. An automatic detection and recognition of malignant clusters of microcalcifications, although very useful for a mass screening of the female population at risk, is very difficult to accomplish because of the small size of the microcalcifications and of the poor quality of the mammographic images. In this paper we propose a novel approach, based on the adoption of a Multiple Expert System (MES). Such a system aggregates several experts, some of which are devoted to classify the single microcalcifications while others are aimed to recognize the malignancy of the cluster considered as a whole. The final classification decision of the system results from the combination of the outputs of the single experts. The approach has been successfully tested on a standard database of 40 mammographic images, publicly available.

1. Introduction

Mammography is a not invasive technique for revealing small tumors, even smaller than 0.5 cm, in women breasts. By suitable indicants appearing in mammographies, a skilled radiologist can detect breast cancer at the earliest possible stage [1]. One of the most used early cancer signs are tiny granule-like deposits of calcium, known as microcalcifications. Microcalcifications appear as small bright spots in a mammogram. Their size ranges from about 0.1 mm to 0.7 mm, with a shape sometimes irregular; they can appear spread all over the breast or grouped in clusters.

It has been experimentally proved that some kinds of microcalcifications are associated to a high probability of cancer [2], and many studies have been concerned to the relations of shape of single and clustered microcalcifications with risky situations [3,4].

In the recent past, the necessity of wide mass screenings increased the need of a computer aided analisys [5]; many pattern recognition methodologies, especially devised for microcalcification detection and classification, have been consequently proposed (e.g. see [6]). Their common objective is that of reducing the false-positive rate, which determines the need of non necessary biopsies, and maintaining, in the same time, the sensitivity of the method. After a mandatory pre-processing step, aimed to improve the quality of mammographies in terms of contrast and noise, all the proposed systems perform a microcalcification detection phase, carried out with rather different approaches [7,8,9]. As regards the classification stage, mainly two different approaches emerge [4,10]: some methods are based on classification of single microcalcifications, while other ones analyze the entire cluster of microcalcifications.

In this paper we propose a novel approach, based on the adoption of a Multiple Expert System (MES) [11] for the recognition of malignant clusters. Such a system aggregates experts based on the classification of single microcalcifications and experts using, on the contrary, features describing the whole cluster. A final stage takes the classification decision on the basis of the responses of each single expert. The approach has been experimented with a standard database of mammographies.

2. The system architecture

Methods based on the recognition of single microcalcifications are affected by the fact that these are difficult to extract from the background and often appear distorted in their shape. This is a quite unresolvable problem as X-ray systems provide digital mammographies

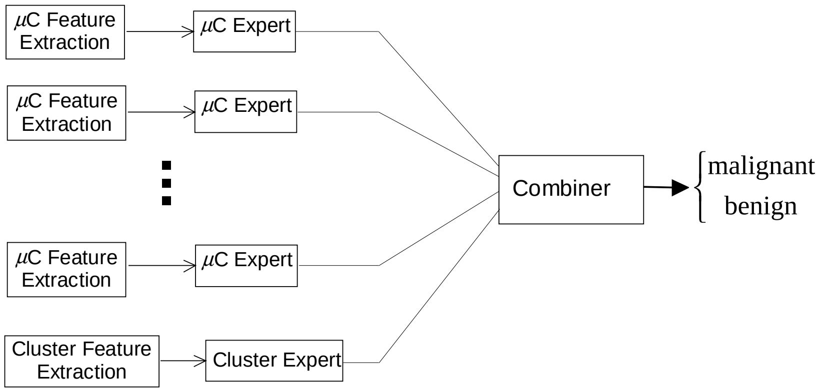

Fig.1: The overall architecture of the proposed system.

at resolutions for which microcalcifications are only few pixels large. So, contrast problems have dramatic consequences on the shape of the microcalcifications.

On the other hand, methods based on the classification of the entire cluster appear to be more robust with respect to shape distortions of single microcalcifications, but they are generally not very reliable when the cluster is weakly described; i.e. when the number of microcalcifications forming the cluster is low.

For this reason, the proposed multi-expert system combines both an expert for classifying single microcalcification (**μC-Expert**) and one for classifying the cluster as a whole (**Cluster Expert**). To decide about the malignancy of a cluster, each microcalcification belonging to it is classified by using the μC-Expert, while the cluster, considered as a whole, is classified by the Cluster Expert. The final response is given by combining the classification decisions coming from all the used experts (see fig. 1).

3. The experts

Each expert is made of a descriptor, which characterizes the input pattern through a set of features, and of a classifier which assigns the pattern to one of the possible classes.

The classifiers adopted to build both the μC-Expert and the Cluster Expert are Multi-Layer Perceptrons (MLPs), trained with the Back Propagation algorithm, with 25 hidden neurons and 2 output neurons, associated to the benign and the malignant class.

The description used by the μC-Expert employs mostly features already proposed in the literature. They refer both to shape properties of the microcalcifications (see table 1) and to the global texture in the area of interest containing the microcalcification (see table 2).

Compactness of the microcalcification
Roughness of the border
Average contrast microcalcification/background
Average local density

Tab. 1: The shape features for the microcalcifications.

For a detailed definition of the adopted features, see [4,6].

Energy in the Area of Interest
Energy in the background
Average luminance
Standard deviation of the luminance
Entropy of the 1st order histogram
Energy of the 2nd order histogram
Contrast of the 2nd order histogram
Entropy of the 2nd order histogram

Tab. 2: The texture features for the microcalcifications.

The set of features describing the cluster mainly take into account its shape and the distribution of the mass within the cluster. It is worth noting that the mass of a microcalcification can be estimated by summing the luminance of the pixels belonging to it. The luminance, in fact, is correlated with the density of total mass crossed by the X-ray beam.

Ellipticity of the cluster
Mass density of the cluster
Average mass of the microcalcifications
Standard deviation of the masses of the microcalcifications
Average distance between microcalcifications and the center of mass of the cluster
Standard deviation of distance between microcalcifications and center of mass

Tab. 3: The cluster features.

4. The combination scheme

Once the experts to combine have been chosen, the main point in the implementation of a MES is the definition of the combining rule most suitable for the application at hand. One of the simplest combining rule, namely the "Majority Voting" [12,13], assigns the input sample the class for which a relative or absolute majority of experts agree, and rejects the sample in the case in which two or more classes receive the same number of votes. More sophisticated combining rules are based on heuristic approaches, like ranking, while others are founded on more formalized theoretical bases, like those based on the Dempster-Shafer evidence theory or on statistical methods [12].

The most suitable combining scheme for the considered case is the "Weighted Voting" [11,14]. According to this rule, the "vote" (i.e. the output) of each expert is weighted by the estimated reliability associated to the expert; all the votes are collected and the input sample is assigned to the class for which the sum of the votes is the highest.

With reference to the application at hand, this rule allows us to weight the outputs coming from the μC Experts in a different way with respect to those coming from the Cluster Expert, so as to take into greater account the more reliable expert. Since the number of microcalcifications in the clusters is variable, we cannot simply combine the output of the Cluster Expert with those of the μC Experts, which, in this case, would be incorrectly weighted in dependence of their number. For this aim, we have considered a two-stage combining scheme: in the first stage (μC aggregation) the results coming from the μC Experts are aggregated and weighted, thus obtaining two confidence degrees about the malignancy (M_m) or benignancy (M_b) of the cluster. This result is then combined with the output of the Cluster Expert, which provides two analogous confidence degrees (C_m and C_b).

For the μC aggregation, two different solutions have been considered. With the first one, the confidence degrees about the malignancy (O_m) and the benignancy (O_b), are evaluated through the average of the outputs of the μC Experts:

$$O_m = \frac{1}{N_{\mu C}} \sum_{i=1}^{N_{\mu C}} O_m^{(i)} \qquad O_b = \frac{1}{N_{\mu C}} \sum_{i=1}^{N_{\mu C}} O_b^{(i)}$$

where $N_{\mu C}$ is the total number of microcalcifications and $O_b^{(i)}$ ($O_m^{(i)}$) is the output value for the benignancy (malignancy) class provided by the i-th μC-Expert. The second aggregation scheme evaluates two analogous confidence degrees (N_m, N_b) by simply counting the number of the microcalcifications classified as malignant (benign) and normalizing these values with respect to the total number of microcalcifications in the cluster. From experiments we have noted that N_m and N_b are generally quite robust with a high number of microcalcifications in the cluster, their reliability sensibly decreases with few microcalcifications, because even only one wrong classification can give rise to imprecise estimates. As a consequence, the values for (M_m, M_b) should be evaluated by combining both (O_m, O_b) and (N_m, N_b), and weighting the two contributions in a different way, on the basis of the number of microcalcifications present within the cluster.

The number of microcalcifications can give useful information also about how to combine the confidence degrees coming from the μC-Experts together with the confidence degrees (C_m, C_b), provided by the Cluster Expert. In fact, the features used to describe the shape and the mass distribution characteristics of the cluster are not very stable with a low number of microcalcifications. Moreover, when $N_{\mu C}$ is high, N_m and N_b appear to be more reliable than the outputs of the Cluster Expert. To this aim, the final combination is realized by means of dynamic weights, evaluated as functions of $N_{\mu C}$, which allow us to emphasize the parameters most reliable for a given number of microcalcifications. The analytical expression devised for the combination is:

$$V_m = \alpha(N_{\mu C}) \cdot O_m + \beta(N_{\mu C}) \cdot N_m + \gamma \cdot C_m$$
$$V_b = \alpha(N_{\mu C}) \cdot O_b + \beta(N_{\mu C}) \cdot N_b + \gamma \cdot C_b$$

where V_m (V_b) is the final vote for the malignancy (benignancy) of the cluster, while $\alpha(N_{\mu C})$ and $\beta(N_{\mu C})$ are the weight functions; the weights for C_m and C_b are both fixed to a costant γ. Typical trends for the weight functions are shown in fig. 2: their parameters are evaluated by means of an optimization phase performed by using a set of clusters representative of the particular domain.

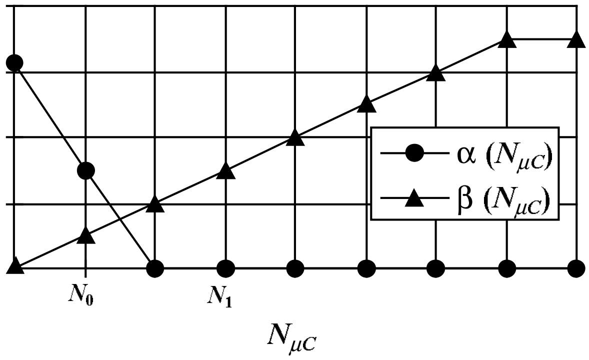

Fig. 2: The qualitative trends of the weight functions.

It is easy to note from fig. 2 that the highest weight is given to the pair (O_m, O_b) for $N_{\mu C} < N_0$ and to the pair (N_m, N_b) for $N_{\mu C} > N_0$; thus, in these two intervals, the contribution coming from the for μC-Experts is prevailing. Conversely, the Cluster Expert succeeds in the range $[N_0, N_1]$.

5. Experimental results

For testing our approach we have used a public database (available at the site http://figment.csee.usf.edu/) of 40 mammographies containing 102 clusters (72 malignant clusters and 30 benign ones), with 1792 malignant and 331 benign microcalcifications. 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. All images have size 2048 by 2048 and use 12 bits (2 bytes) per pixel of gray level information. Some preprocessing was performed to convert the images to a 8 bit/pixel format using an adaptive noise equalisation described in [15].

This database represents a severe test bed, since the size of the microcalcifications is typically very small. Moreover, the low number of clusters (specially of benign clusters) makes very difficult the learning of experts based on neural networks. For this reason, we have adopted a leave-one-out approach: for each cluster in the database, we employ the remaining clusters to constitute the training sets for learning the μC-Expert and the Cluster Expert. A successive optimization phase provides the parameters of the weight functions. The classification is finally performed on the cluster extracted.

A first analysis of the performance obtained with our approach has been carried out by assigning the cluster to the class with the highest vote between V_m and V_b. The relative results are shown in table 4: the first column provides the recognition rate (in percentage) obtained by considering only the Cluster Expert; the successive two columns report the results obtained by the μC-Expert with the two possible aggregations, while the last column contains the results given by the whole MES.

It is worth noting that the recognition rate exhibited by the MES on the malignant clusters is lightly smaller than the best result obtained by one expert, while there is a significant improvement in the recognition of benign clusters and thus a significant decrease of false positives.

Another goal of our experiments was to measure the ability of the MES to recognize malignant clusters, avoiding erroneous recognition of benign clusters as malignant, which could cause a significant increase of non productive biopsy examinations. Two figures used to this aim are the *True Positive* (*TP*) and the *False Positive* (*FP*) rates defined as:

$$TP = \frac{malignant\ clusters\ correctly\ classified}{total\ malignant\ clusters}$$

$$FP = \frac{benign\ clusters\ incorrectly\ classified}{total\ benign\ clusters}$$

	Cluster Expert	μC-Expert (O_b,O_m)	μC-Expert (N_b,N_m)	**MES**
Malignant Clusters	60.00 %	75.74 %	75.37 %	75.37 %
Benign Clusters	70.00 %	67.5 %	60.00 %	73.50 %

Tab. 4: The recognition rates obtained by each expert and by the whole MES.

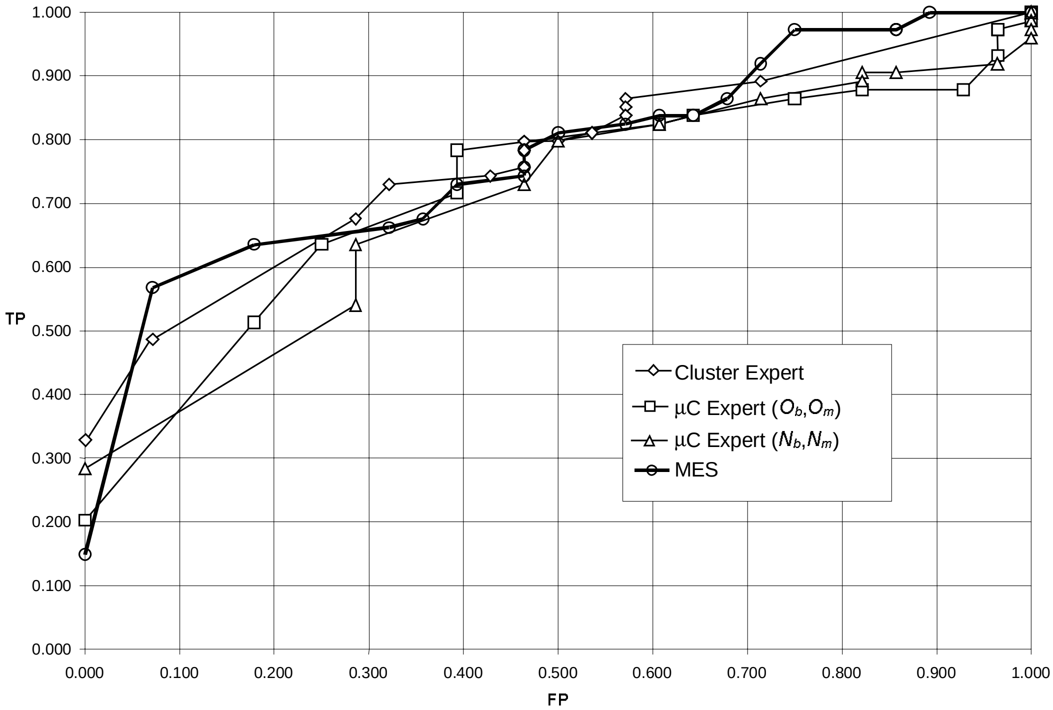

Fig. 3: The ROC curves of the single experts and of the whole MES.

In order to evaluate which error tradeoffs the MES can attain, we have normalized the output values of the MES, so as to vary between 0 and 1, and imposed on the normalized value $\overline{V_m}$ a threshold in the range [0,1]. For each value assumed by the threshold, we consider only the clusters with $\overline{V_m}$ higher than the threshold and compute *TP* and *FP*. The resulting curve provides a *Receiver Operating Characteristic* (*ROC*) graph, which is extensively used for visualizing and analyzing the accuracy of diagnostic systems [16]. ROC graph reports *TP* on the Y axis and *FP* on the X axis; informally, the nearer the curve to the upper right point of the diagram, the better the performance obtained (higher *TP* and/or lower *FP*).

Fig. 3 shows the ROC curves evaluated for each single expert and for the whole MES. It is possible to note that the MES performs better both for low and for high values of *FP*, while it has quite the same performance of the single experts for medium values of *FP*. An immediate way to globally compare the MES with respect the single experts is to measure the area under the respective ROC curves. In the ideal case (expert with TP=1 for each $FP \in [0,1]$), this measure is 1; in real situations, the more the area approaches 1, the better the diagnostic system. Table 5 shows the results obtained. Also in this case it is possible to note that the performances exhibited by the MES are better than each single expert.

	Cluster Expert	μC-Expert (O_b,O_m)	μC-Expert (N_b,N_m)	**MES**
Area under the ROC curve	0.73	0.68	0.66	0.78

Tab. 5: The results obtained by each expert and by the whole MES in terms of area under the ROC curve.

References

[1] R. Mc Clelland "Screening for Breast Cancer: Opportunities, status and challenges" in *Recent Results in Cancer Research*, vol. 119, S. Brunner and B. Langfeldt eds., Springer-Verlag, Berlin, pp. 29-38, 1990.

[2] M. Lanyi, Diagnosis and differential diagnosis of breast calcifications, Springer-Verlag, New York, 1986.

[3] B. Zheng, W. Qian, and L. P. Clarke, "Digital Mammography: Mixed Feature Neural Network with Spectral Entropy Decision for Detection of Microcalcifications", *IEEE Trans. on Medical Imaging*, vol. 15, no. 5, pp. 589-597, 1996.

[4] L. Shen, R.M. Rangayyan, J.E.L. Desautels, "Application of Shape Analysis to Mammographic calcifications", *IEEE Trans. on Medical Imaging*, vol. 13, no. 2, pp. 234-253, 1994.

[5] E.D. Pisano, F. Shtern "Image processing and Computer Aided Diagnosis in digital mammography: A clinical perspective", *Int. Journal of Pattern Recognition and Artificial Intelligence*, vol. 7, no. 6, pp. 1493-1503, 1993.

[6] A.P. Dhawan, Y. Chitre, C. Kaiser-Bonasso, M. Moskowitz, "Analysis of Mammographic Microcalcifications Using Gray-Level Image Structure Features", *IEEE Trans. on Medical Imaging*, vol. 15, no. 3, pp. 246-259, 1996.

[7] K. Woods, C. Doss, K. Bowyer, J. Solka, C. Priebe, W. P. Kegelmeyer Jr., "Recognition Techniques for Detection of Microcalcifications in Mammography", *Int. Journal of Pattern Recognition and Artificial Intelligence*, vol. 7, no. 6, pp. 1417-1436, 1993.

[8] T.C. Wang, N. Karayiannis, "Detection of Microcalcifications in Digital Mammograms Using Wavelets", *IEEE Trans. on Medical Imaging*, vol. 17, no. 4, pp. 498-509, 1998.

[9] H. Cheng, Y.M. Lui, R.I. Freimanis, "A Novel Approach to Microcalcification Detection Using Fuzzy Logic Technique", *IEEE Trans. on Medical Imaging*, vol. 17, no. 3, pp. 442-451, 1998.

[10] Y. Chitre, A.P. Dhawan, M. Moskowitz, Artificial Neural Network Based Classification of Mammographic Microcalcifications Using Image Structure Features", *Int. Journal of Pattern Recognition and Artificial Intelligence*, vol. 7, no. 6, pp. 1377-1401, 1993.

[11] L. Xu, A. Krzyzak, C.Y. Suen, "Method of Combining Multiple Classifier Systems and Their Application to Handwritten Numeral Recognition", *IEEE Trans. on Systems, Man and Cybernetics*, vol. 22, no. 3, pp 418-435, 1992.

[12] B. Ackermann, H. Bunke, "Combination of Classifiers on the Decision Level for Face Recognition", Technical Report IAM-96-002, Institut für Informatik und angewandte Mathematik, Universität Bern, 1996.

[13] L. Lam, C.Y. Suen, "Increasing Experts for Majority Vote in OCR: Theoretical Considerations and Strategies", *Proc. of the 4th International Workshop on Frontiers in Handwriting Recognition*, pp. 245-254, 1994.

[14] T.K Ho, J.J. Hull, S. Shrihari, "Decision Combination in Multiple Classifier Systems", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol 16, no.1, pp.66-75, 1994.

[15] N. Karssemeijer, "Adaptive Noise Equalization and Recognition of Microcalcification Clusters in Mammograms", *Int. Journal of Pattern Recognition and Artificial Intelligence*, vol. 7, no. 6, pp. 1357-1376, 1993.

[16] J. Swets, "Measuring the Accuracy of Diagnostic Systems", *Science*, vol. 240, pp. 1285-1293, 1988.