Microcalcification Detection using a Kernel Bayes Classifier

B. Caputo (*,+),E. La Torre (+), G. E. Gigante (+)

(*) Smith-Kettlewell Eye Research Institute
2318 Fillmore Street, San Francisco
94115 California, USA
e-mail: caputo@ski.org
(+) University of Rome 'La Sapienza',
CISB- Physics Department,
Piazza A. Moro 3, 00161, Rome, Italy
e-mail: {b.caputo, elatorre, g.gigante}@caspur.it

Abstract. Mammography associated with clinical breast examination is the only effective method for mass breast screening. Microcalcifications are one of the primary signs for early detection of breast cancer. In this paper we propose a new kernel method for classification of difficult-to-diagnose regions in mammographic images. It consists of a novel class of Markov Random Fields, using techniques developed within the context of statistical mechanics. This method is used for the classification of positive Region of Interest (ROI's) containing clustered microcalcifications and negative ROI's containing normal tissue. We benchmarked the new proposed method with a nearest neighbor classifier and with an artificial neural network, widely used in literature for computer-aided diagnosis. We obtained the best performance using the novel approach.

1 Introduction

The usage of machine learning algorithm for the development of computer--assisted diagnosis system is a well developed area of research [9, 16]. The diagnostic process is a very complicated procedure, the outcome of which depends on numerous factors. Some of these factors relate to the procedure itself and others are associated with characteristics of the human visual system. In the situation in which all technical factors are at a high quality level, the human factor becomes more important. The perception of details and the recognition of the meaning of these details are the weakest links in the biomedical image interpretation process; perceptual and cognitive problems occur that will result in loss of information and a lower quality of the resulting diagnosis. Thus the importance of developing computer-assisted diagnosis systems which make use of knowledge developed within the machine learning community.

A challenging application problem is the detection of microcalcifications in X-ray mammograms. Screen-films mammography associated with clinical breast examination and breast self examination is widely recognized as the only effective

imaging modality for early detection of breast cancer in women [14, 6]. However, the interpretation of X-ray mammograms is very difficult because of the small differences in the image densities of various breast tissues, particularly for dense breast. The interpretation of mammograms by radiologists is performed by a visual examination of films for the presence of abnormalities that indicate cancerous changes. Computerized analysis might be of significant value to improve the true-positive rate of breast cancer detection. Among the early indicators of breast cancer, microcalcifications are one of the primary signs. They are tiny granule-like depositum of calcium, and the presence of clustered microcalcifications in X-ray mammograms is considered a basic marker for the early detection of breast cancer, especially for individual microcalcification with diameters up to about 0.7 mm and with an average diameter of 0.3 mm [14, 6].

Computerized image analysis methods have been used for the identification of circumscribed masses, classification of suspicious areas and classification of microcalcifications using conventional methods [8], [13] and using expert systems [8]. In the actual interpretation of mammographic microcalcifications, the gray-level values defining local structures in the microcalcification clusters play a significant role [6]. It has been demonstrated in clinical studies described in [6], that the grouping of microcalcification regions, in order to define the shape of the cluster, is highly dependent on the gray-level-based structure and texture of the image. Texture information plays an important role in image analysis and understanding, with potential applications in remote sensing, quality control, and medical diagnosis. Texture is one of the important characteristics used in identifying an object or a region of interest (ROI) in an image [5]. A vast literature has addressed this topic in the last decades [8, 9, 5, 16, 13, 3]. Most of work has been devoted to the search of key textural features for the representation of significant information. The classification step is generally performed using Artificial Neural Networks (ANN).

In this paper we focus the attention on the choice of the classification algorithm rather than on the choice of the textural features. We propose to use Spin Glass -Markov Random Fields (SG-MRF, [4]) for microcalcification detection. SG-MRF is a fully connected MRF which integrates results of statistical physics of disordered systems [1] with Gibbs probability distributions via non linear kernel mapping [12]. SG-MRF have shown to be very effective for many visual applications such as appearance-based object recognition, texture classification and so forth [4]. Here we apply the very same strategy for microcalcification detection. We represent each Region Of Interest (ROI) using a shape histogram representation [11], then we model the SG-MRF on the histogram bins. This probabilistic model is used to classify ROI into positive ROIs containing microcalcifications and negative ROIs containing normal tissue. The classification step is performed using a Maximum A Posteriori (MAP) probability classifier. We compare SG-MRF's performance with that obtained using a Nearest Neighbor Classifier (NNC) and an ANN given by a three-layer perceptron, with back-propagation learning algorithm. In all the experiments we performed, SG-MRF obtained the best performances. To the best of our knowledge, these are the first

experiments on microcalcification detection that show a better performance of a given method (in this case SG-MRF) with respect to a multi-layer perceptron classifier.

The paper is organized as follows: Section 2 describes the algorithm used for the feature extraction step. Section 3 reviews basic concepts of ANN, and Section 4 summarizes the SG-MRF model and how it can be employed for classification purposes in a MAP classifier. Section 5 presents experimental results; the paper concludes with a summary discussion.

2 Feature Extraction: Multidimensional Receptive Field Histograms

Multidimensional receptive Field Histograms (MFH) were proposed by Schiele [11] in order to extend the color histogram approach of Swain and Ballard [15]. The main idea is to calculate multidimensional histograms of the response of a vector of receptive fields. A MFH is determined once we chose the local property measurements (i.e., the receptive field functions), which determine the dimensions of the histogram, and the resolution of each axis. On the basis of the results reported in [4], we chose to use in this research work two local characteristics based on Gaussian derivatives:

$$D_x = -\frac{x}{\sigma^2}G(x,y), \qquad D_y = -\frac{x}{\sigma^2}G(x,y), \tag{1}$$

where

$$G(x,y) = exp\left(-\frac{x^2+y^2}{2\sigma^2}\right) \tag{2}$$

is the Gaussian distribution. The parameter σ explicitly determines the scale of the filter, and it will be specified later.

3 Artificial Neural Networks

Also referred to as connectionist architectures, parallel distibuted processing, and neuromorphic systems, an artificial neural network (ANN) is an information--processing paradigm inspired by the way the densely interconnected, parallel structure of the mammalian brain processes information. Artificial neural networks are collections of mathematical models that emulate some of the observed properties of biological nervous systems and draw on the analogies of adaptive biological learning. The key element of the ANN paradigm is the novel structure of the information processing system. It is composed of a large number of highly interconnected processing elements that are analogous to neurons and are tied together with weighted connections that are analogous to synapses [7], [10], [2]. A weight w_{ij} (coupling strength) characterizes the interconnections between any two neurons i and j. The input to each neuron is a weighted sum of the outputs incoming from the connected neurons. Each neuron operates on the input

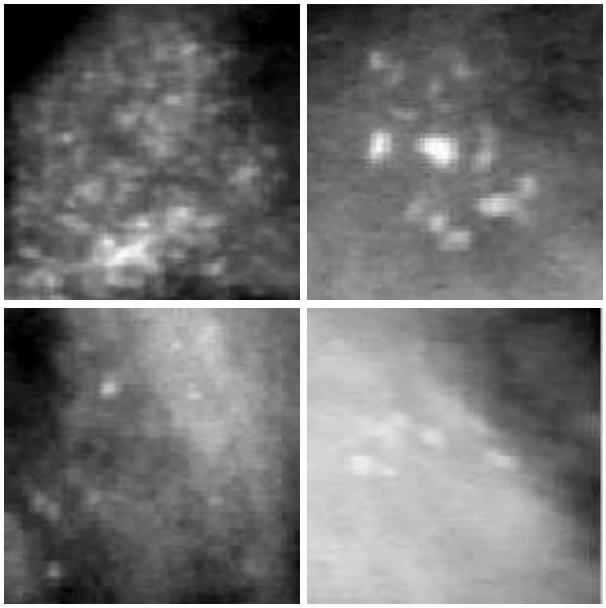

Fig. 1. Four examples of ROIs containing microcalcifications.

signal using his activation function f and produces the output response. The typical activation functions are linear, threshold and sigmoid [10], [2]. Normally the neurons are organized in an architecture with input nodes, interfacing the neural network and the external world, output nodes, producing the network's responses, and hidden nodes, having the task of correlating and building up an "internal representation" of the analyzed problem. Network's capacity and performance depends on the number of neurons, on the activation functions used, and on the neurons' interconnections. Another important attribute of artificial neural networks is that they can efficiently learn nonlinear mappings through examples contained in a training set, and use the learned mapping for complex decision making [10], [2].

Although ANNs have been around since the late 1950's, it wasn't until the mid-1980's that algorithms became sophisticated enough for general applications. Today ANNs are being applied to an increasing number of real-world problems of considerable complexity. They are good pattern recognition engines and robust classifiers, with the ability to generalize in making decisions about imprecise input data. The advantage of ANN lies in their resilience against distortions in the input data and their capability of learning. They are often good at solving problems that are too complex for conventional technologies (e.g., problems that do not have an algorithmic solution or for which an algorithmic solution is too complex to be found) and are often well suited to problems that people are good at solving, but for which traditional methods are not.

4 Spin Glass-Markov Random Fields

Consider a visual category Ω_j and a set of k observations $\{\mathbf{x}^1 \dots \mathbf{x}^k\}, \mathbf{x} \in \Re^m$, that we consider random samples from the underlying, unknown, probability distribution $P(\mathbf{x})$ defined on $\Re^m$. Consider also $\mathcal{K}$ different visual categories $\Omega_j, j = \{1, \dots \mathcal{K}\}$. Given an observation $\hat{\mathbf{x}}$, our goal is to classify $\hat{\mathbf{x}}$ as a sample from Ω_{j^*}, one of the Ω_j object classes. Using a Maximum A Posteriori (MAP) criterion we have

$$j^* = \underset{j}{\operatorname{argmax}} P(\Omega_\kappa|\mathbf{x}) = \underset{j}{\operatorname{argmax}}\{P(\mathbf{x}|\Omega_j)P(\Omega_j)\};$$

using Bayes rule, where $P(\mathbf{x}|\Omega_j)$ are the Likelihood Functions (LFs) and $P(\Omega_j)$ are the prior probabilities of the classes. Assuming that $P(\Omega_j)$ are constant, the Bayes classifier simplifies to

$$j^* = \underset{j}{\operatorname{argmax}} P(\mathbf{x}|\Omega_j).$$

A possible strategy for modeling the parametric form of the probability distribution is to use Gibbs distributions within a MRF framework.

Spin Glass-Markov Random Fields (SG-MRFs) [4] are a new class of MRFs which connect SG-like energy functions (mainly the Hopfield one [1]) with Gibbs distributions via a non linear kernel mapping. The resulting model overcomes many difficulties related to the design of fully connected MRFs, and enables to use the power of kernels in a probabilistic framework. The SG-MRF probability distribution is given by

$$P_{SG-MRF}(\mathbf{x}|\Omega_j) = \frac{1}{Z}\exp\left[-E_{SG-MRF}(\mathbf{x}|\Omega_j)\right], \tag{3}$$

$$Z = \sum_{\{\mathbf{x}\}} \exp\left[-E_{SG-MRF}(\mathbf{x}|\Omega_j)\right],$$

with

$$E_{SG-MRF} = -\sum_{\mu=1}^{p_j}\left[K(\mathbf{x}, \tilde{\mathbf{x}}^{(\mu)})\right]^2, \tag{4}$$

where the function $K(\mathbf{x}, \tilde{\mathbf{x}}^\mu)$ is a Generalized Gaussian kernel [12]:

$$K(\mathbf{x}, \mathbf{y}) = \exp\{-\rho d_{a,b}(\mathbf{x}, \mathbf{y})\}, d_{a,b}(\mathbf{x}, \mathbf{y}) = \sum_i |x_i^a - y_i^a|^b.$$

$\{\tilde{\mathbf{x}}^\mu\}_{\mu=1}^{p_j}, j \in [1, \mathcal{K}]$ are a set of vectors selected (according to a chosen ansatz, [4]) from the training data that we call *prototypes*. The number of prototypes per class must be finite, and they must satisfy the condition:

$$K(\tilde{\mathbf{x}}^i, \tilde{\mathbf{x}}^k) = 0, \tag{5}$$

for all $i, k = 1, \dots p_j, i \neq k$ and $j = 0, \dots \mathcal{K}$. The interested reader can find a detailed discussion regarding the derivation and properties of SG-MRF in [4].

5 Experimental Results

We tested the performance of SG-MRFs for microcalcifications detection on a database of 81 images produced by the "Centro per la Cura e la prevenzione dei Tumori" of the University of Rome "La Sapienza"; each image was digitized from film using a CCD camera operating at a spatial resolution of 604×575 pixels for image; the pixel rate was of $11,5 MHz$, and the pixel size of $10\mu m \times 15\mu m$. From the 81 images, 152 Region of Interest (ROI) were selected by expert radiologists, each of 128×128 pixels. Among the selected 152 ROIs, 112 were positive and 40 were negative; four different ROIs are shown in Figure 1. We used as training set 59 images representing positive ROIs, and 33 images representing negative ROIs. The rest of the database was used as a test set. In a preprocessing step, each extracted ROI was stretched to the normalized gray-level range of 0-255 [5]. Features were extracted using a Multidimensional receptive Field Histogram (MFH) representation [11][1], that has been already used successfully combined with SG-MRF [4]. We used 2D MFH, with filters given by Gaussian derivatives along x and y directions as described in Section 2 and with $\sigma = 1.0$; resolution for histogram axis of 16 bins. For the classification step, we used SG-MRF in the MAP-MRF framework described in Section 3. For the choice of prototypes, we made a naive ansatz [4], which means that all training views are taken as prototypes, and the ρ in the Gaussian kernel is learned so to satisfy condition (5). The performance of SG-MRF was compared with a Nearest Neighbor Classifier (NNC) and an ANN. More precisely, we used a three-layer perceptron, with backpropagation learning algorithm. the textural features extracted by means of MFH histograms are used as the input signals of the input layer. there is a single output node for classification into positive or negative ROI. The performance of SG-MRF was evaluated as the kernel parameters a, b varied. The performance of the ANN was evaluated as the number of neurons in the hidden layer varied.

NNC	SG-MRF				ANN			
88.33	$a = 1, b = 1.5$	88.33	$a = 0.5, b = 1.5$	93.33	$S_1 = 1$	83.33	$S_1 = 4$	83.33
	$a = 1, b = 1$	90.00	$a = 0.5, b = 1$	93.33	$S_1 = 2$	83.33	$S_1 = 5$	83.33
	$a = 1, b = 0.5$	93.33	$a = 0.5, b = 0.5$	93.33	$S_1 = 3$	85.00	$S_1 = 6$	81.67

Table 1. Classification results for NNC, ANN and SG-MRF. S_1 represents the number of neurons in the hidden layer for ANN.

Classification results are reported in Table 1. The best recognition rate, corresponding to 93.33%, was obtained using SG-MRF. This result is obtained with more than one combination of kernel parameters (see Table 1). NNC gives a best performance of 88.33%, corresponding to a +5% less with respect to the recognition rate obtained using SG-MRF. ANN gives a best performance of 85.00%,

[1] We gratefully thank B. Schiele who allowed us to use his software for the computation of MFH.

corresponding to the worst recognition rate obtained on this database. It corresponds to a 3.33% loss with respect to the recognition rate obtained using NNC, and to an impressive 8.33% loss with respect to the recognition rate obtained using SG-MRF.

These results show the effectiveness of SG-MRF for microcalcification detection. At the same time, they show that, in the building of algorithms for computer-assisted diagnosis, ANN cannot be considered a priori the optimal choice for the classification step. To the best of our knowledge, these are the first experiments that reports of a comparative study on microcalcification detection with respect to the kind of classifier employed.

6 Summary

In this paper we presented a new probabilistic approach for microcalcification detection. It applies a new kernel method, Spin Glass-Markov Random Fields, that already proved to be very effective for many visual applications such as object recognition and scene classification [4]. The method is benchmarked with a NNC and an ANN on the same feature representation, obtaining respectively an impressive +5 % and +8.33% recognition rate. This experimental result shows the effectiveness of the proposed approach.

This work can be extended in many ways. First, the performance of SG-MRF can be improved choosing a different set of kernel parameters, and a different representation. Second, we plan to benchmark this approach with other classifiers such as support vector machines. Finally, we intend to compare the performance of SG-MRF and the aforementioned classifiers using different set of features. Future work will concentrate in these directions.

Acknowledgments We would like to thank Prof. V. Virno and the staff of the Radiology department of the "Centro per la cura e la Prevenzione dei Tumori" of the University of Rome "La Sapienza". B. C. has been supported by the Foundation BLANCEFLOR Boncompagni-Ludovisi.

References

1. D. J. Amit, *"Modeling Brain Function"*, Cambridge University Press, 1989.
2. C. M. Bishop, *Neural Networks for Pattern Recognition*, Claredon Press - Oxford, 1995.
3. B. Caputo, G. E. Gigante, "Digital Mammography: a Weak Continuity Texture Representation for Detection of Microcalcifications", Proc. of SPIE Medical Imaging 2001, February 17-22, VOL 4322, PP1705-1716, San Diego, (CA), USA, 2001.
4. B. Caputo, H. Niemann, "From Markov Random Fields to Associative Memories and Back: Spin Glass Markov Random Fields", SCTV2001.
5. A. K. Jain, "*Fundamental of digital image processing*", Prentice Hall, Englewood Cliffs, 1989.

6. M. Lanyi, "*Diagnosis and Differential Diagnosis of Breast Calcifications*", New York: Springer-Verlag, 1986.
7. R. P. Lippmann, "An introduction to computing with neural nets", *IEEE ASSP Magazine*, pp. 4-22, April 1987.
8. S. Morio and S. Kawahara et al., "Expert system for early detection of cancer of the breast", *Comp. Biol. Med.*, vol. 19, no. 5, pp. 295-305, 1989.
9. Nishikawa RM, Wolverton DE, Schmidt RA, Papaioannou J, "Radiologists' ability to discriminate computer-detected true and false positives, from an automated scheme for the detection of clustered microcalcifications on digital mammograms", *Proc SPIE 3036*: 198-204, 1997.
10. D. E. Rumelhart and C.R. Rosemberg, *Parallel Distributed Processing*, the MIT Press, Cambridge MA, 1986.
11. B. Schiele, J. L. Crowley, "Recognition without correspondence using multidimensional receptive field histograms",*IJCV*, 36 (1), pp. 31- 52, 2000.
12. B. Schölkopf, A. J. Smola, *Learning with kernels*, 2001, the MIT Press.
13. L.Shen, R. M. Rangayyan and J. E. L. Desautels, "Application of shape analysis to mammographic calcifications", *IEEE Trans. Med Imag.*, vol. 13, no. 2, pp. 263-274, 1994.
14. E. A. Sickles and D. B. Kopans, "Mammographic screening for women aged 40 to 49 years: the primary practitioner's dilemma", *Anna. Intern. Med.*, vol. 122, no. 7, pp. 534-538, 1995.
15. M. Swain, D. Ballard, "color Indexing", *IJCV*, 7, pp 11-32, 1991.
16. Zhang W, Doi K, Giger ML, Nishikawa RM, Schmidt RA, "An improved shift-invariant artificial neural network for computerized detection of clustered microcalcifications in digital mammograms". *Med Phys*, 23: 595-601, 1996.