

Preliminary results of a featureless CAD system on FFDM images

Renato Campanini¹, Enrico Angelini¹, Danilo Dongiovanni¹, Emiro Iampieri¹,

NicoLanconelli¹, Claude Mair-Noack³, Matteo Masotti¹, Giuseppe Palermo¹,

Matteo Roffilli^{1,2}, Gianni Saguatti⁴, Omar Schiaratura²

¹ Department of Physics, University of Bologna and INFN – Bologna, Italy

² Computer Science, University of Bologna, Italy

³ Stadtspital Triemli, Zurich, Switzerland

⁴ Ospedale Maggiore, Bologna, Italy

Abstract. A novel featureless approach to the detection of masses and microcalcifications has been adopted, based on a Support Vector Machine (SVM) classifier. This method does not rely on any feature extraction task; on the contrary, the algorithm automatically learns to detect the lesions by the examples presented to it during the training phase. Our technique includes a pre-selection step, in which we select the intra-breast areas that will be analyzed. Those regions are then provided to an SVM classifier, trained to recognize suspect masses or microcalcifications. The CAD performance have been already assessed on digitized mammogram freely available on the net (DDSM USF and Nijmegen databases). In this paper we are going to test the CAD scheme on digital images coming from Giotto Image MD FFDM unit, a mammography system based on an amorphous Selenium detector. Images have a spatial resolution equal to 85 μm and 13 bit gray-level resolution and have been collected at two different sites: Maggiore Hospital in Bologna (Italy) and Triemli Hospital in Zurich (Switzerland). Preliminary results are presented on a database gathered at these hospitals. The CAD system marked in the FFDM images 19 cancers out of 23, with a false-positive rate of 0.9 marks per image.

1. Introduction

Masses and clustered microcalcifications are the most common lesions associated with the presence of breast carcinomas. The automatic detection of these lesions can be hampered by the wide diversity of their shape, size and subtlety. The lesions can vary considerably in optical density, shape, position, size and characteristics at the edge. Therefore, it is difficult to identify morphological, directional or structural quantities that can characterize the lesions sought at any scales and any modalities of occurrence. As a consequence, for a CAD system it is very demanding to detect lesions of various types. The reason is that detection methods often rely on a feature extraction step: here, the lesions are isolated by means of a set of characteristics. Due to the great variety of the lesions, it is extremely difficult to get a common set of features effective for every kind of lesion.

In this paper, we present a detection system, which does not rely on any feature extraction step. Considering the complexity of the class of objects to be detected, considering that said objects frequently present characteristics that are very similar to the environment which surround them, and considering the objective difficulty of modeling this class of objects with few measurable quantities, in the approach proposed herein no modeling has been used. On the contrary, the algorithm automatically learns to detect the lesions by the examples presented to it. In this way, there is no *a priori* knowledge provided by the trainer: the only thing the system needs is a set of positive examples and a set of negative examples. Basically, we consider lesion detection as a two-class pattern recognition problem. The detection scheme codifies the image with a wavelet overcomplete representation; the great amount of information handled by the algorithm is classified by means of a Support Vector Machine (SVM) classifier, a learning machine based on a well-founded statistical theory. Given the ability of SVMs to handle multidimensional spaces, at the same time maintaining a good generalization capacity, the possibility of eliminating or limiting the feature extraction step

for a classification task has emerged. SVMs have already been applied to breast cancer detection methods. Basically, the SVM was used for reducing false-positive signals, by means of extracted image features: in the detection of microcalcifications in mammograms (Bazzani et al. 2001), and in the diagnosis of breast ultrasonography images (Chang et al. 2003). A featureless approach based on SVM for the detection of lesions in mammograms has been investigated for the first time by our group (Campanini et al. 2002, Campanini et al. 2004). A similar approach has been adopted for the detection of microcalcifications by El-Naqa et al. (El-Naqa et al. 2002).

In this study, we validated our detection scheme with images coming from two FFDM units: one located at “Stadtsptal Triemli”, Zurich, Switzerland and another one at “Ospedale Maggiore”, Bologna, Italy. The two systems are “Giotto IMAGE MD” produced by IMS, Italy (Albanese et al. 2002). They are based on amorphous Selenium flat panel digital detector manufactured by ANRAD Corporation, Canada. The active area of the imager is $17.4\text{ cm} \times 23.9\text{ cm}$ with a pixel pitch of $85\text{ }\mu\text{m}$; images have 2048×2816 pixels with 13 bit gray-level resolution.

2. Materials and methods

2.1. Detection scheme

Our detection algorithm has been trained to detect both clustered microcalcifications and masses. Figure 1 shows a chart of our detection scheme. The rationale of the two branches is that in both cases the detection is performed without using any extracted image feature. Our algorithm automatically extracts the needed information during the training.

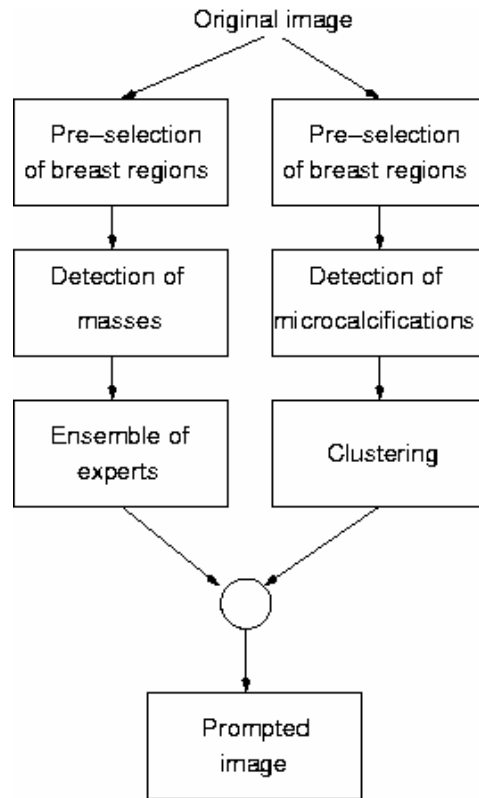


Figure 1. Flow-chart of the featureless detection method.

The first step of the masses detection algorithm consists in a pre-selection of the suspect regions within the breast. This is achieved by means of adaptive local gray-level thresholding. All the pixels selected are then analyzed by an ensemble of three different experts. Each expert is able to carry out a multiscale detection, in order to find out masses with size ranging from 3 mm to 35 mm. The searching performed by each expert is based on the SVM classification of the overcomplete wavelet representation of all the crops centered on the pixels selected in the first step. Finally, a region is marked as suspect mass thanks to a *voting* strategy of the committee of the three experts. Basically, a region is considered suspect only if at least two of the three experts detect that region. More details of the masses featureless detection algorithm can be found in Campanini et al. 2004.

The first step of the microcalcification detection method consists in a pre-selection of the regions containing bright spots. This is achieved by means of a gaussianity test calculated on a linear-filtered image (Bazzani et al. 2001). Pixels that pass the gaussianity test are then

provided to a detector similar to the experts used for the masses. Here, a wavelet overcomplete representation of the crops centered on the points extracted in the first step is obtained. After that, the crops are judged as positive or not, by using an SVM classifier. The main difference of the featureless detection between masses and microcalcifications is that in the first case a multiscale searching is used, whereas in the second case crops of fixed size are considered. The single adjacent pixels classified as suspect are grouped together and clusterized, if more than two signals in a 1 cm^2 area are detected.

Finally, signals discovered by the masses and clustered microcalcifications detectors are joined by means of a logical OR operator, and a maximum predetermined number of marks are presented as the final result. Signals are ranked by means of their distance from the separating hyperplane (Campanini et al. 2004).

2.2. Image dataset

A former version of our detection scheme has already been tested on digitized images freely available on the net. In particular, the masses featureless algorithm has been recently evaluated on the DDSM-USF database (Campanini et al. 2004). A previous method for the detection of clustered microcalcifications based on extracted image features has been tested on the Nijmegen database (Bazzani et al. 2001). For the first time we are going to evaluate the entire featureless detection algorithm on digital FFDM images. The dataset used consists of about 750 images coming from two different “Giotto Image MD” FFDM systems. Images have a pixel size equal to 85 μm and a gray-level resolution of 13 bits; they have been collected both in the course of the clinical evaluation of the FFDM system and subsequently during the regular clinical examinations. Each case is relative to one patient and comprises four projections (two cranio-caudal and two medio-lateral views). The database includes 672 normal images (without lesions) and 88 images with at least one lesion, such as tumor

opacities or clustered microcalcifications. The location of the lesions have been marked by expert radiologists and collected together with the images. Digital mammograms were always available in four projections per patient. The dataset has been divided into two subsets: a training set and a test set, used for the setting and the testing of the detection algorithm, respectively. The training set consists of 46 cancer images and 52 normal ones, whereas the test is performed on 42 cancer images and 620 images without lesions.

3. Results

Since the number of marked images available for the CAD training is small, we decided to use the training previously accomplished with digitized images coming from USF DDSM database (Heat et al. 2000). In this work, we applied a sigmoidal Look-Up-Table (LUT), for transforming the histogram of the FFDM images. In order to close the optimal LUT, we trained the system with 44 positive crops of lesions from our FFDM dataset and about 4000 negative crops from normal images. In this way, we find out the best LUT, which maps the FFDM images histogram into the histogram relative to the images used for the already achieved CAD training. That has allowed us to exploit the very large number of images from DDSM database, for testing the CAD on the still small FFDM dataset available. The CAD system marked in the FFDM images 19 cancers out of 23, with a false-positive rate of 0.9 marks per image. Figure 2 shows the FROC curve computed on the test images. The sensitivity has been calculated as the number of positive patients correctly detected over the total number of positive patients. The false-positive rate has been computed on the normal images.

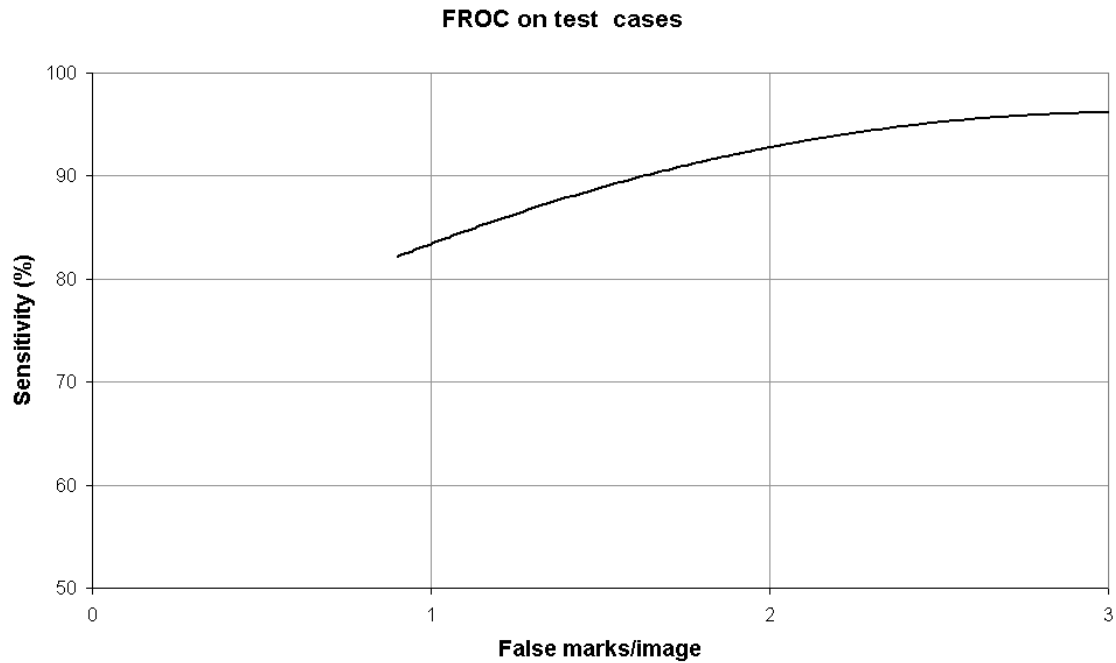


Figure 2. FROC curve of the featureless algorithm on FFDM test cases.

Future work has to be done, in order to enlarge the dataset of FFDM images. In this way, it will be possible to train new experts directly on those images, allowing an improvement of the CAD results and a more precise determination of its performance.

References

- Albanese A., A. Bevilacqua, R. Campanini, E. Iampieri, N. Lanconelli, M. Rossi, D. Romani, V. Salomoni, and P. Vignoli. 2002. Characterization of an FFDM unit based on a-Se direct conversion detector. *Digital Mammography: IWDM2002, 6th International Workshop on Digital Mammography*. Springer: 69-71.
- Bazzani A., A. Bevilacqua, D. Bollini, R. Brancaccio, R. Campanini, N. Lanconelli, A. Riccardi, and D. Romani. 2001. An SVM classifier to separate false signals from microcalcifications in digital mammograms. *Phys. Med. Biol.* 46:1651-1663.

Campanini R., D. Dongiovanni, E. Iampieri, N. Lanconelli, M. Masotti, G. Palermo, A. Riccardi, and M. Roffilli. 2004. A novel featureless approach to mass detection in digital mammograms based on Support Vector Machines. *Phys. Med. Biol.* 49:961-975.

Campanini R., A. Bazzani, A. Bevilacqua, D. Bollini, D. Dongiovanni, E. Iampieri, N. Lanconelli, A. Riccardi, M. Roffilli, and R. Tazzoli. 2002. A novel approach to mass detection in digital mammography based on Support Vector Machines (SVM). *Digital Mammography: IWDM2002, 6th International Workshop on Digital Mammography*. Springer: 399-401.

Chang R.F., W.J. Wu, W.K. Moon, Y.H. Chou, and D.R. Chen. 2003. Support vector machines for diagnosis of breast tumors on US images. *Acad. Radiol.* 10:189–197.

El-Naqa I., Y. Yang, M.N. Wernick, N.P. Galatsanos, and R.M. Nishikawa. 2002. A Support Vector Machine approach for detection of microcalcifications. *IEEE Trans. Med. Imag.* 21:1552-1563.

Heat M., K. Bowyer, D. Kopans, R. Moore, and P. Kegelmeyer. 2000. The digital database for screening mammography. *Digital Mammography: IWDM2000 5th Int. Workshop on Digital Mammography*. Medical Physics Publishing: 212-218.