Multiple-Instance Learning for Breast Cancer Detection in Mammograms

Rubén Sánchez de la Rosa, Mathieu Lamard, Guy Cazuguel, Gouenou Coatrieux, Michel Cozic and Gwenolé Quellec

Abstract— **This paper describes an experimental computer-aided detection and diagnosis system for breast cancer, the most common form of cancer among women, using mammography. The system relies on the Multiple-Instance Learning (MIL) paradigm, which has proven useful for medical decision support in previous works from our team. In the proposed framework, the breasts are first partitioned adaptively into regions. Then, either textural features, or features derived from the detection of masses and microcalcifications, are extracted from each region. Finally, feature vectors extracted from each region are combined using an MIL algorithm (Citation k-NN or mi-Graph), in order to recognize "normal" mammography examinations or to categorize examinations as "normal", "benign" or "cancer". An accuracy of 91.1% (respectively 62.1%) was achieved for normality recognition (respectively three-class categorization) in a subset of 720 mammograms from the DDSM dataset. The paper also discusses future improvements, that will make the most of the MIL paradigm, in order to improve "benign" versus "cancer" discrimination in particular.**

I. Introduction

Breast cancer is the most common form of cancer among women worldwide. Currently, mammography is the most effective tool for early detection of breast cancer, which increases the survivability of patients. In order to help radiologists interpret mammograms, many computer-aided detection and diagnosis systems have been developed in recent years [1]. Computer-Aided Detection (CADe) systems are developed to help the radiologist in locating the abnormal areas in mammograms. Computer-Aided Diagnosis (CADx) systems, which sometimes rely on CADe systems, are developed to diagnose cancer based on the analysis of these anomalies. CADx involves detecting the breast in images and detecting anomalies within the breast, masses and microcalcifications in particular; CADe involves characterizing these anomalies and building a cancer detector based on these characterizations [2].

In recent studies, we have shown the relevance of the Multiple-Instance learning (MIL) paradigm for both medical image and medical video analysis applications [3], [4]. This paper investigates MIL within the context of mammography CADe and CADx. MIL is a useful paradigm to solve image analysis problems where visual features are extracted locally in images, but the labels used for supervision are assigned to images as a whole (e.g. "normal", "benign", "cancer"). Even though publicly available datasets, such as DDSM [5], come with manual segmentations, which may be used for supervision, that property is useful to train CADx algorithms on large, real-life datasets where no manual segmentations are available. It should be noted that three studies have already investigated the use of MIL in this context: MIL was used either for mass detection [6], [7] or for microcalcification detection [8], i.e. within the scope of a CADe system. In this paper, in line with our works of diabetic retinopathy screening [3], the goal is to screen cancer, possibly using mass and microcalcification detections, i.e. to develop a CADx system or a CADe + CADx system. Initial results in the DDSM dataset are reported and future developments are discussed.

R. Sánchez de la Rosa, G. Cazuguel and G. Coatrieux are with INSTITUT TELECOM; TELECOM Bretagne; UEB; Dpt ITI, Brest, F-29200 France
M. Lamard, G. Cazuguel, G. Coatrieux and G. Quellec are with Inserm, UMR 1101, Brest F-29200, France gwenole.quellec@inserm.fr
M. Lamard is with Univ Bretagne Occidentale, Brest, F-29200 France
M. Cozic is with Medecom, Plougastel-Daoulas, F-29470 France

II. Multiple-Instance Learning

Multiple-instance learning (MIL) generalizes supervised learning. In supervised learning, each instance is described by a feature vector and is associated with a class label. In MIL, each instance is also described by a feature vector. But in that case, the class label is not associated with an instance: it is associated with a *bags of instances*, containing an arbitrary large number of instances [9]. In image processing, this representation is useful when images are divided into regions and a feature vector is extracted from each region, regarded as an instance. If no manual segmentation is available for images, then only the class label of the image, regarded as a bag of instances, is known.

Many algorithms exist for solving MIL problems: two of them are presented below and evaluated in this paper.

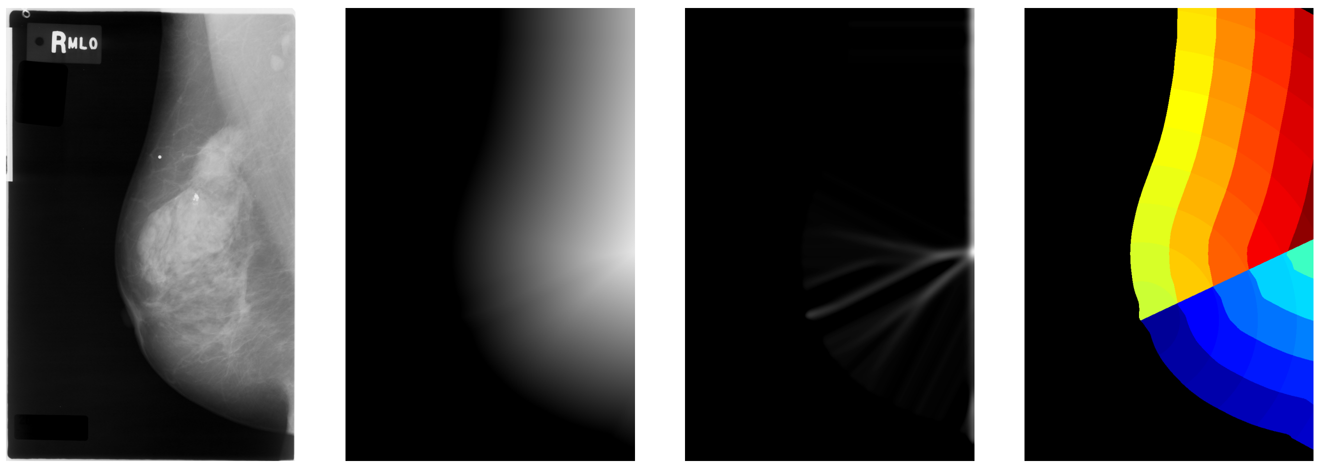
A. Citation k-NN

One of the most popular MIL algorithms is Citation k-NN [10], which generalizes the even more popular k nearest neighbors (k-NN) classifier. In the MIL adaptation, the minimum Hausdorff distance, defined as the shortest distance between any two instances from each bag, is used as the bag-level distance metric. Citation k-NN considers not only the nearest neighbors of a given bag (known as references), but also the bags that count it as one of their neighbors (known as citers), hence the algorithm's name.

B. mi-Graph

A recent benchmarking study [11], focusing on two computer-aided diagnosis problems, found mi-Graph [12] to

978-1-4244-9270-1/15/$31.00 ©2015 IEEE



(a) original optical density image (b) distance transform (c) distance transform ridges (d) breast's main axis and region segmentation

Fig. 1: Breast segmentation and region definition

be the best performing MIL algorithm for these applications [12]. In mi-Graph, each bag is represented by a graph of similarities among its instances. To compare two bags, the inter-bag and intra-bag instance similarities are used to define a bag-level kernel function. This kernel function puts the emphasis on unusual instances, i.e. instances which are dissimilar from other instances within their bag. This kernel function is used to train a support-vector machine.

C. Multiple-instance learning for Breast Cancer Screening

In order to apply these algorithms to mammograms, each image must be divided into regions and a feature vector should be extracted from each region. A region definition framework is presented in section III. Two types of features are investigated in section IV: texture features and features deriving from anomaly detections. It should be noted that in the particular case of mammography screening, the class label is not assigned to single images, but rather to mammography examinations, consisting of four images (two view per breast). So a bag of instances groups together regions from four images.

III. Partitioning the Breast into Regions

MIL algorithms usually rely on a partition of images into rectangular regions. Because a large portion of mammograms is irrelevant (black background, labels, etc.), we propose to define regions within the breast only (see section III-B), and to define regions adaptively, so that regions fit the shape of the breast, which implies detecting the breast edges (see section III-B) and the nipple (see section III-C).

A. Preprocessing Images (see Fig. 1 (a))

Before processing images, the first step is to map the gray levels in input images to optical densities [5], which are then encoded on 12 bits. Therefore, it is possible to process images acquired with various scanners, and possibly digitized with various digitizers. For faster breast segmentation and nipple detection, images are downsampled by a factor of four (the original definition is used for feature extraction in section IV).

B. Segmenting the Breast

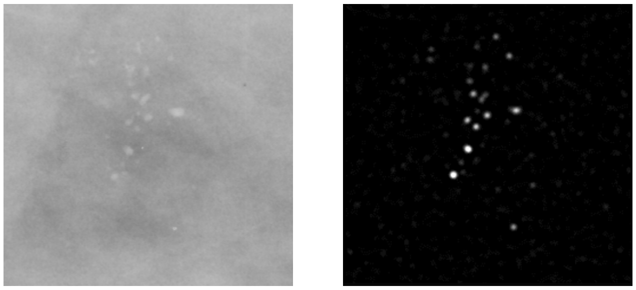
Bright and sharp image areas are detected using a threshold on image intensities and a threshold on intensity derivatives. The reason why we detect sharp areas is that intensity is decreasing near the breast edges, so these edges usually are not bright, but they are sharp. To make sure noise is not enhanced with the derivative filter, the image is smoothed beforehand using a median filter. Holes, that may appear between sharp and bright areas in the resulting binary image are filled. The breast is defined as the largest connected component.

C. Finding the Main Axis and the Nipple (see Fig. 1 (b)-(c))

Once the breast is segmented, the distance from each pixel in the breast to the nearest pixel on the breast edges (excluding the image border itself), is computed using the Maurer distance transform [13] (see Fig. 1 (b)). Ridges of this distance map are then detected using a large Laplacian filter (see Fig. 1 (c)). It can be observed that the main ridge runs orthogonally to the breast edges and intersects those edges very close to the nipple. Bright ridge pixels are then fitted with a line, referred to as the main axis. A robust estimation of the nipple is given by the intersection of this line and of the breast edges.

D. Defining Regions (see Fig. 1 (d))

Image pixels are grouped in regions with respect to their distance to the breast edges and to their distance to the nipple. The distance transform defined above is used to partition the breast into uniform intervals of distances to the breast edges. The breast is also partitioned into uniform intervals of Euclidean distances to the nipple; the same interval width is used for both distances. Finally, the breast is partitioned



(a) region of interest (b) local anomaly probabilities

Fig. 2: Microcalcification segmentation.

into two halves, based on the main axis. Regions are defined as intersections of these partitions.

IV. Feature Extraction

Two types of features are extracted from each of region defined above: either texture features (see section IV-A) or features derived from anomaly detections (see section IV-B).

A. Texture Features

In a first scenario, several popular texture features are extracted from a given region: Haralick features [14], run-length features [15], a histogram of local binary patterns [16], and a gray level histogram. All these features are then concatenated. The dimensionality of the resulting feature vector is very high. To avoid overfitting issues, its dimensionality is not reduced by a supervised feature selection approach, but rather by a principal component analysis. This compressed feature vector is used to describe the region.

B. Features Derived from Anomaly Detection

In a second scenario, two types of anomalies are detected in images: masses and microcalcifications (MCs). A recent algorithm was used for mass detection: a topographic approach by Hong et al. [17]. A novel solution was used for MC detection: a point of interest detector by Loy et al. [18], called Fast Radial Symmetry Transform (FRST). To our best knowledge, this detector, which looks for roundish objects in images, has never been applied to MC detection before. A detection example is reported in Fig. 2. Once these anomalies are detected, a histogram of local anomaly probabilities is computed for each region and each type of anomalies. The feature vector describing a given region is defined as the concatenation of both histograms.

V. Experiments on the DDSM Dataset

A. The DDSM Dataset

The Massachusetts General Hospital, the University of South Florida and the Sandia National laboratories have collected, for research purposes, a dataset of mammography examinations, called the Digital Database for Screening Mammography (DDSM) [5]. This dataset consists of 2620 mammography examinations, each examination including two images of each breast, along with some associated patient information. Each mammogram comes with a manual segmentation roughly delineating masses and microcalcifications.

TABLE I: Classification performance (accuracy)

No. of classes	texture-based recognition		anomaly-based recognition	
	Citation k-NN	mi-Graph	Citation k-NN	mi-Graph
2	58.3%	71.6%	79.0%	**91.1%**
3	26.7%	57.7%	57.0%	**62.1%**

B. Performance of Anomaly Detection in the DDSM Dataset

Based on these manual segmentations, the performance of mass and microcalcification detection at object level, in the full DDSM dataset, is reported in Fig. 3. These results were obtained for our own implementations of the topological mass detector [17] and of the FRST detector [18].

C. Performance of MIL-Based Screening in a Subset of DDSM

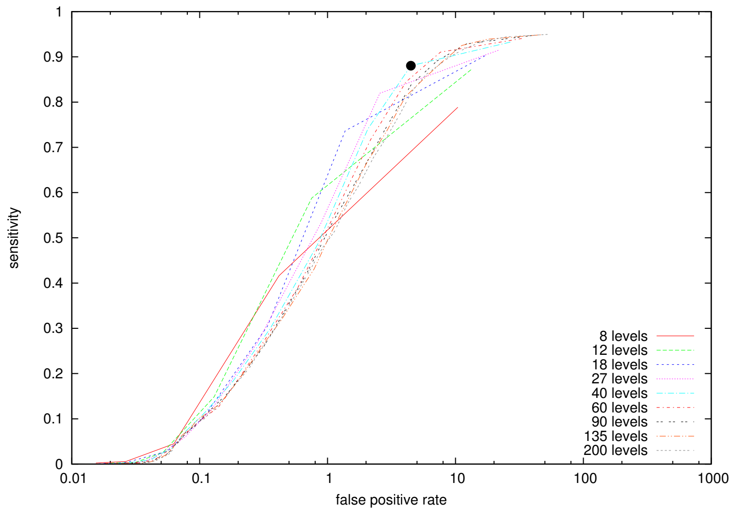
A random subset of 180 mammography examinations, containing 720 mammograms altogether, was used to validate breast segmentation qualitatively and to validate the MIL principle quantitatively. A uniform distribution of class labels was used: 60 “normal”, 60 “benign” and 60 “cancer” cases.

1) Breast Segmentation and Nipple Detection: from a visual inspection, we observed that breast segmentation and nipple detection were successful in these 720 mammograms, with the exception of minor errors that can be observed, for instance, at the bottom of Fig. 1. This validates the robustness of our main axis detection algorithm (see section III-C) used to define regions adaptively.

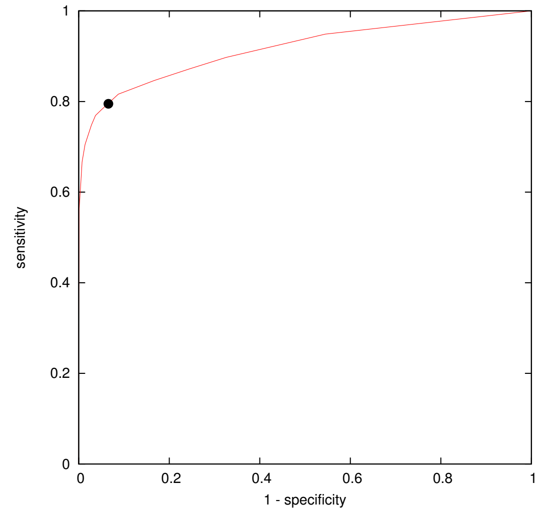
2) Classification Performance: to evaluate MIL-based classification, two scenarios were investigated: a two-class and a three-class scenario. In the three-class scenario, the three classes are: “normal”, “benign” and “cancer”. In the two-class scenario, the two classes are: “normal” and “benign or cancer” (i.e. “abnormal”). The classification performance in terms of accuracy, obtained by 5-fold cross-validation, is reported in table I.

VI. Discussion

This preliminary study has shown that Multiple-Instance Learning (MIL) can potentially help detect cancer in mammography examinations. Three observations can be made: 1) the use of anomaly detection is more relevant than the use of texture in this context, 2) mi-Graph is more relevant than k-NN Citation, which confirms a recent benchmarking study [11], and 3) whereas the performance of anomaly detection is good (accuracy = 91.1%), the performance of “cancer” versus “benign” classification needs to be improved. Regarding the first observation, we believe the combination of both cues will improve performance. In particular, a two-stage procedure will be investigated. In this procedure, 1) candidate regions will be selected through anomaly-based MIL to maximize “normal” versus “abnormal” classification and 2) “cancer” versus “benign” classification will be performed



(a) mass detection



(b) microcalcification detection

Fig. 3: Evaluating the performance of mass and microcalcification segmentation in the DDSM dataset. Fig. (a) reports the detection performance of masses against manual segmentations, for different numbers of topographic levels, the main algorithm parameter [17]. Fig. (b) reports the ability to discriminate regions with and without microcalcifications, according to the manual segmentations, based on the average local anomaly probability (see Fig. 2 (b)). Black circles indicate the optimal settings used in the MIL framework.

through texture-based MIL using the selected regions only, in order to simplify the classification problem. Moreover, it should be noted that this framework does not make the most of the region definition we have proposed: we plan to use the region coordinates (with respect to the breast edges and to the nipple) to push performance forward in future works. To conclude, in line with our works on MIL for medical decision support, we have reported a preliminary computer-aided diagnosis system for mammography screening that will be improved in future works.

VII. Acknowledgments

This work was supported in part by a grant from the Brittany Region, through the ID2Santé program (DéCaRICA), and in part by a grant from the French *Agence Nationale de la Recherche* (ANR, LabCom SePEMeD).

References

[1] A. Jalalian, S. B. Mashohor, H. R. Mahmud, M. I. Saripan, A. R. Ramli, and B. Karasfi, "Computer-aided detection/diagnosis of breast cancer in mammography and ultrasound: a review," *Clin Imaging*, vol. 37, no. 3, pp. 420–6, May–June 2013.

[2] K. Ganesan, U. R. Acharya, C. K. Chua, L. C. Min, K. T. Abraham, and K. H. Ng, "Computer-aided breast cancer detection using mammograms: a review," *IEEE Rev Biomed Eng*, vol. 6, pp. 77–98, 2013.

[3] G. Quellec, M. Lamard, M. D. Abràmoff, E. Decencière, B. Lay, A. Erginay, B. Cochener, and G. Cazuguel, "A multiple-instance learning framework for diabetic retinopathy screening," *Med Image Anal*, vol. 16, no. 6, pp. 1228–40, August 2012.

[4] G. Quellec, M. Lamard, B. Cochener, and G. Cazuguel, "Real-time task recognition in cataract surgery videos using adaptive spatiotemporal polynomials," *IEEE Trans Med Imaging*, 2015, in press.

[5] M. Heath, K. Bowyer, D. Kopans, R. Moore, and W. P. Kegelmeyer, "The digital database for screening mammography," in *Proc Int Workshop Digital Mammography*, M. J. Yaffe, Ed., 2001, pp. 212–8.

[6] B. Krishnapuram, J. Stoeckel, V. Raykar, B. Rao, P. Bamberger, E. Ratner, N. Merlet, I. Stainvas, M. Abramov, and A. Manevitch, "Multiple-instance learning improves CAD detection of masses in digital mammography," in *Digital Mammography – LNCS*, vol. 5116, 2008, pp. 350–357.

[7] P. Lu, W. Liu, W. Xu, L. Li, B. Zheng, J. Zhang, and L. Zhang, "Multi-instance learning for mass retrieval in digitized mammograms," in *Proc SPIE Medical Imaging*, vol. 8315, 2012, p. 831523.

[8] C. Li, K. M. Lam, L. Zhang, C. Hui, and S. Zhang, "Mammogram microcalcification cluster detection by locating key instances in a multi-instance learning framework," in *Proc IEEE ICSPCC*, 2012, pp. 175–9.

[9] J. Amores, "Multiple instance classification: review, taxonomy and comparative study," *Artif Intell*, vol. 201, pp. 81–105, August 2013.

[10] J. Wang and J.-D. Zucker, "Solving the multiple-instance problem: a lazy learning approach," in *Proc Int Conf Mach Learn*, 2000, pp. 1119–25.

[11] M. Kandemir and F. A. Hamprecht, "Computer-aided diagnosis from weak supervision: A benchmarking study," *Comput Med Imaging Graph*, 2014, in press.

[12] Z.-H. Zhou, Y.-Y. Sun, and Y.-F. Li, "Multi-instance learning by treating instances as non-I.I.D. samples," in *Proc ICML*, 2009, pp. 1249–56.

[13] C. R. Maurer Jr., R. Qi, and V. Raghavan, "A linear time algorithm for computing exact Euclidean distance transforms of binary images in arbitrary dimensions," *IEEE Trans Pattern Anal Mach Intell*, vol. 25, no. 2, pp. 265–70, February 2003.

[14] R. M. Haralick, K. Shanmugam, and I. D. Dinstein, "Textural features for image classification," *IEEE Trans Syst Man Cybern*, vol. SMC-3, no. 6, pp. 610–21, November 1973.

[15] M. M. Galloway, "Texture analysis using gray level run lengths," *Comput Graph Image Process*, vol. 4, no. 2, pp. 172–9, June 1975.

[16] T. Ojala, M. Pietikinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern Recognit*, vol. 29, no. 1, pp. 51–9, January 1996.

[17] B. W. Hong and B. S. Sohn, "Segmentation of regions of interest in mammograms in a topographic approach," *IEEE Trans Inf Technol Biomed*, vol. 14, no. 1, pp. 129–39, January 2010.

[18] G. Loy and A. Zelinsky, "Fast radial symmetry for detecting points of interest," *IEEE Trans Pattern Anal Mach Intell*, vol. 25, no. 8, pp. 959–73, August 2003.