Medical Image Analysis 18 (2014) 241–252

Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

Learning from unbalanced data: A cascade-based approach for detecting clustered microcalcifications

A. Bria [a,*], N. Karssemeijer [b], F. Tortorella [a]

[a] Department of Electrical and Information Engineering, University of Cassino and L.M., Via Di Biasio 43, 03043 Cassino (FR), Italy
[b] Diagnostic Image Analysis Group, Radboud University Nijmegen Medical Centre, P.O. Box 9102, 6500 HC Nijmegen, The Netherlands

ARTICLE INFO

Article history:
Received 26 June 2013
Received in revised form 18 October 2013
Accepted 31 October 2013
Available online 12 November 2013

Keywords:
Computer aided detection
Unbalanced data
Clustered microcalcifications
Mammography

ABSTRACT

Finding abnormalities in diagnostic images is a difficult task even for expert radiologists because the normal tissue locations largely outnumber those with suspicious signs which may thus be missed or incorrectly interpreted. For the same reason the design of a Computer-Aided Detection (CADe) system is very complex because the large predominance of normal samples in the training data may hamper the ability of the classifier to recognize the abnormalities on the images. In this paper we present a novel approach for computer-aided detection which faces the class imbalance with a cascade of boosting classifiers where each node is trained by a learning algorithm based on ranking instead of classification error. Such approach is used to design a system (*CasCADe*) for the automated detection of clustered microcalcifications (μCs), which is a severely unbalanced classification problem because of the vast majority of image locations where no μC is present. The proposed approach was evaluated with a dataset of 1599 full-field digital mammograms from 560 cases and compared favorably with the Hologic R2CAD ImageChecker, one of the most widespread commercial CADe systems. In particular, at the same lesion sensitivity of R2CAD (90%) on biopsy proven malignant cases, CasCADe and R2CAD detected 0.13 and 0.21 false positives per image (FPpi), respectively (p-value = 0.09), whereas at the same FPpi of R2CAD (0.21), CasCADe and R2CAD detected 93% and 90% of true lesions respectively (p-value = 0.11) thus showing that CasCADe can compete with high-end CADe commercial systems.

1. Introduction

Clustered microcalcifications (μCs) are one of the most important early indicators of breast cancer since they appear in 30–50% of cases diagnosed by mammographic screenings (Kopans, 2007). However, interpreting screening mammograms is a big challenge even for an expert radiologist since the low prevalence makes finding abnormalities difficult. Birdwell (2009) points out several subjective factors that may lead to a lack of perception or to mistakes in interpretation. Among the established methods to improve radiologist performance, it has been reported that having more than one radiologist or a Computer-Aided Detection (CADe) system improves the detection of cancer in mammograms (Karssemeijer et al., 2009; Eadie et al., 2012). To this end, several commercial CADe systems are nowadays available and their use is widespread among radiologists. However, even though CADe systems show a sensitivity similar to radiologists (Cole et al., 2012), there are still a few hundred false positives for every true positive in a screening setting, which is about two orders of magnitude higher than what the radiologists achieve (Karssemeijer et al., 2009) and this potentially limits the benefit that a CADe system can provide. For this reason, the design of CADe systems for clustered μCs is still an open research field as shown by the recent literature (El Naqa et al., 2002; Wei et al., 2005; Tang et al., 2009; Zhang et al., 2009; Oliver et al., 2010; Jing et al., 2011).

Among the proposed approaches, methods based on supervised learning techniques have received the largest share of research since they can yield powerful binary classifiers able to determine whether a μC is present (positive) or not (negative) at a pixel location. However, such methods have to face two major problems. First, the huge number of pixels to be analyzed (e.g., about 9 million in a digital mammogram) coupled with high-complexity classifiers may cause a computational burden not easy to sustain, especially when hundreds or thousands of images have to be processed. Second, the vast majority of image locations where no μC is present makes detection a severely unbalanced classification problem, where the negative class is several orders of magnitude bigger than the positive class. Generally speaking, binary classifiers trained on highly unbalanced data sets tend to be overwhelmed by the majority class, thus misclassifying the samples belonging to the minority class. This problem is also known as class imbalance and in recent years it has received considerable attention in

* Corresponding author. Tel.: +39 3480339824.
E-mail addresses: a.bria@unicas.it (A. Bria), n.karssemeijer@rad.umcn.nl (N. Karssemeijer), tortorella@unicas.it (F. Tortorella).

1361-8415/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.media.2013.10.014

the machine learning community (e.g., Barandela et al., 2003; Guo et al., 2008) and subsequently in the medical image analysis field (Li et al., 2010). Several approaches focused on μC detection (e.g., Wei et al., 2005; Zhang et al., 2009; Oliver et al., 2010; Marrocco et al., 2010) address class imbalance by randomly selecting a limited set of negative samples so as to obtain approximately the same size for the two classes. Nevertheless, there is no guarantee that the selected subset is actually representative of all the possible negative samples. A different solution is proposed in El Naqa et al. (2002) in which a Support Vector Machine (SVM) is employed with a *Successive Enhancement Learning* (SEL) scheme where the SVM is initially trained with a balanced training set containing a limited number of negative samples. The training is then restarted iteratively by incorporating another N misclassified negative samples from all the available training images. The retraining step is repeated until no more changes are observed in support vectors. In this way the total number of training samples is kept small and balanced at each retraining round, but the final classifier could be very complex since it could contain a very large number of support vectors, thus making the detection phase computationally intense.

In this paper we present *CasCADe*, a multistage system for the automatic detection of clustered μCs on full-field digital mammograms (FFDM), specifically designed to handle efficiently and effectively the computational complexity and the high class imbalance. Even though aimed at the μC detection problem, the approach proposed in this work could be more generally applicable in medical image analysis and especially in other unbalanced problems such as the automated detection of lung nodules in CT (e.g., vanGinneken10 et al., 2010), chest lymph nodes in CT (e.g., Barbu et al., 2012; Feulner et al., 2013), colon polyps in CT colonography (e.g., Van Ravesteijn et al., 2010), and retinal microaneurysms in Digital Color Fundus Photographs (e.g., Niemeijer et al., 2010). The rationale of our approach is to employ an ensemble of ranking-based boosting classifiers connected in series with increasing complexity and specificity like in the cascade face detector proposed by Viola and Jones (2001). The choice of boosting-based classifiers is particularly fitting for unbalanced problems as demonstrated by Galar et al. (2011) who empirically compared the most significant published approaches and showed that for two-class unbalanced problems the best results were obtained by using random undersampling techniques coupled with bagging or boosting ensembles. In our approach each classifier stage is trained with only a part of the negative samples, thus distributing the complexity of the whole problem among the classifiers and alleviating class imbalance at the same time. Nevertheless, the residual imbalance present at each node could produce unsatisfactory results if the learning algorithm used in the node is based on the optimization of a performance measure (such as the empirical error) highly affected by the class distribution skew. This is the case of AdaBoost, the learning algorithm employed in the approach of Viola and Jones. The same authors observe in a successive paper (Viola and Jones, 2002) that AdaBoost minimizes a quantity related to the classification error (and not the number of false negatives) and thus propose a variant aimed at moderating the effects of the class imbalance by introducing an asymmetric weight updating mechanism of the samples in the training set.

The novelty of our approach firstly lies in handling the class imbalance in each node through a boosting algorithm designed to maximize the Area under the ROC curve (AUC). The reason for this choice is that AUC is equivalent to the probability of correct pairwise ranking and thus provides a measure of the predictive ability of the classifier which is robust and insensitive to the class skew (Huang and Ling, 2005). To this end we adopted a reformulation of RankBoost for bipartite ranking problems (Freund et al., 2003), suitably modified to be embedded in a cascade structure. Another difference from the approach of Viola and Jones (2001) is that our cascade-based detector is used not only for μC localization, but also for accurately estimating the outline of a μC, which has been proven to play an important role for the automated differentiation between true positive and false positive detected μCs (Veldkamp and Karssemeijer, 1996). Indeed, after grouping μCs into clusters, we classify them into "abnormal" (true positive) and "normal" (false positive) clustered μCs, the latter including both benign clusters of μCs and erroneously detected clusters. The low prevalence of cancer within a mammographic screening cohort makes also this decision an unbalanced problem and thus we employ again a RankBoost classifier. To this end, we also propose a novel set of features especially aimed at capturing the topological relations between μCs.

The detection performance of CasCADe was evaluated on 1599 full-field digital mammograms from 560 cases obtained in routine screening and compared with the one of the most widespread commercial CADe systems, the Hologic R2CAD ImageChecker. To our knowledge, the scientific literature does not exhibit other CADe systems which have been compared with high-end commercial systems.

2. Method

The CasCADe system consists of a preprocessing stage, an initial detection stage and a classification stage in which the number of false positive detected clusters is reduced. A schematic overview of these stages is given in Fig. 1. Each of these stages is detailed in the following subsections.

2.1. Preprocessing stage: quantum noise equalization

In FFDM the dominant source of noise is quantum noise that is caused by fluctuations in photon fluence at the detector. These fluctuations can be described by a Poisson distribution with standard deviation $\sqrt{\lambda}$, where λ is the average number of detected photons (Beutel et al., 2000; Schie and Karssemeijer, 2008). Since in an FFDM system a linear relationship exists between gray level and exposure, quantum noise standard deviation σ_q can be estimated by (e.g., McLoughlin et al., 2004; Schie and Karssemeijer, 2008):

$$\sigma_q(y) = c\sqrt{y} \tag{1}$$

where c is a noise level parameter to be estimated and y is the pixel intensity.

Actually noise properties vary across the image and thus c should be estimated locally as, for example, proposed in Schie and Karssemeijer (2008). However, the same study reports that the improvement in μC detection performance obtained with a nonuniform noise model is quite limited and thus we adopted an uniform noise model in which c is constant across the image. On this basis, in order to rescale pixel intensities to a scale with uniform noise level, we consider a scale transform $y' = T(y)$ that satisfies the following differential equation:

$$dT(y) = \frac{\alpha}{\sigma_q(y)}dy \tag{2}$$

where α is the noise level on the transformed scale and the factor $\sigma_q(y)^{-1}$ eliminates the dependency of the differential dy on the noise variation. Coupling Eq. (2) with Eq. (1) and with border conditions, we obtain the following Cauchy problem:

$$\begin{cases} \dfrac{dT(y)}{dy} = \dfrac{\alpha}{c\sqrt{y}} \\ T(0) = 0 \\ T(y_{max}) = T_{max} \end{cases} \tag{3}$$