Addressing Image Variability While Learning Classifiers for Detecting Clusters of Micro-calcifications

Glenn Fung[1], Balaji Krishnapuram[1], Nicolas Merlet[2], Eli Ratner[2], Philippe Bamberger[2], Jonathan Stoeckel[3], and R. Bharat Rao[1]

[1] Computer Aided Diagnosis & Therapy group, Siemens Medical Solutions USA, Inc., 51 Valley Stream Pkwy, Malvern, PA-19355, USA
[2] Siemens Computer Aided Diagnosis, 8 Hartoum St, Beck building, Jerusalem 91450, Israel
[3] Siemens Information Systems Ltd, 84 Keonics, Hosur Road, Bangalore 590100, India

Abstract. Computer aided detection systems for mammography typically use standard classification algorithms from machine learning for detecting lesions. However, these general purpose learning algorithms make implicit assumptions that are commonly violated in CAD problems. We propose a new ensemble algorithm that explicitly accounts for the small fraction of outlier images which tend to produce a large number of false positives. A bootstrapping procedure is used to ensure that the candidates from these outlier images do not skew the statistical properties of the training samples. Experimental studies on the detection of clusters of micro-calcifications indicate that the proposed method significantly outperforms a state-of-the-art general purpose method for designing classifiers (SVM), in terms of FROC curves on a hold out test set.

1 Introduction

In *computer aided diagnosis* (CAD) applications the goal is to detect structures of interest to physicians in medical images: *e.g.* to identify potentially malignant lesions in mammography. In an almost universal paradigm, this problem is addressed by a 3 stage system: identification of potentially unhealthy candidate *regions of interest* (ROI) from a medical image, computation of descriptive features for each candidate, and classification of each candidate (*e.g.* normal or diseased) based on its features.

This paper focusses on automatic algorithms for designing (*i.e.* learning) pattern classifiers for the third stage. Automatic learning algorithms are an important part of the modern methodology for efficiently designing computer aided diagnostic products. Besides improving the diagnostic accuracy, these technologies greatly reduce the time required to develop algorithms that act as “second readers”.

In the context of computer aided mammography, many standard algorithms (*e.g.* support vector machines, Back-propagation for Neural Nets, Kernel Fisher Discriminants) have been used to learn classifiers for detecting malignant lesions in computer aided mammography [1, 2, 3]. However, these general purpose learning methods make implicit assumptions that are commonly violated in CAD applications, often resulting in sub-optimal prediction accuracy for the classifiers that they learn. For example, these methods almost universally assume that the training samples are *independently* drawn from an *identical*—albeit unobservable—underlying distribution (i.i.d. assumption).

Susan M. Astley et al. (Eds.): IWDM 2006, LNCS 4046, pp. 84–91, 2006.

We propose a new ensemble algorithm that is designed to improve the classification accuracy. This algorithm explicitly accounts for the fact that a set of outlier images tend to produce a large number of false (true) positives in the training set used to learn classifiers, whereas many other images only contribute relatively few positive (negative) samples each. A bootstrapping procedure is used to ensure that the candidates from these outlier images do not skew the statistical properties of the training set.

When we learnt a classifier using a standard state-of-the-art method—support vector machine (SVM)—for detecting clusters of micro-calcs, the resulting system performed (generalized) poorly on a hold out set of test samples, in terms of per-image sensitivity & per-patient sensitivity. By contrast, the proposed methods significantly improved the ROC curves, especially in the operating region of interest (around 0.2 FP per image).

The rest of the paper is organized as follows. Section 2 highlights some of the assumptions that underly almost all algorithms for learning pattern classifiers, and indicates why some of them may be inappropriate for CAD. Based on this analysis, Section 3 develops a novel method for learning classifiers that detect clusters of microcalcifications. Experimental results are provided in Section 4. We conclude with a discussion of the broader applicability of the proposed algorithm and some ideas for future extensions in Section 5.

2 Common Assumptions While Learning Pattern Classifiers

2.1 Creation of the Training Data

During the design of a CAD system, considerable human intervention and domain knowledge engineering is employed in the first two stages of a CAD system for (a) candidate generation (CG): identifying all potentially suspicious regions in a candidate generation stage with very high sensitivity, and (b) feature-extraction: description of each such region quantitatively using a set of medically relevant features. For example quantitative measurements based on texture, shape, intensity and contrast and other such characteristics may be used to characterize any region of interest (ROI). Subsequently, for learning the classifier to be used in the third stage, a training dataset is created by obtaining features which describe each candidate ROI in the training images, and class labels are assigned to them based upon the overlap and/or distance from any radiologist-marked (diseased) region.

2.2 Characteristic Properties of the Data

A few important characteristics of the data are relevant for designing classifiers that generalize well. First, there is a form of stochastic dependence between the labeling errors of a group of candidates, all of which are spatially proximate to the same radiologist mark. Further, the features used to describe spatially adjacent or overlapping samples are also highly correlated. As a result, both the labels and the features for the training samples from an image tend to be highly correlated: the inter sample correlation is particularly high for spatially adjacent candidates.

Second, some types of biological or image structures tend to be identified much more often by CG algorithms in the form of many spatially adjacent candidates. This