Support Vector Machines for Candidate Nodules Classification

Paola Campadelli [a] , Elena Casiraghi [a] and Giorgio Valentini [a]

[a] *DSI - Dipartimento di Scienze dell'Informazione*
Università degli Studi di Milano, Italy.

Abstract

Image processing techniques have proved to be effective for the improvement of radiologists' diagnosis of lung nodules. In this paper we present a computerized system aimed at lung nodules detection; it employs two different multi-scale schemes to identify the lung field and then extract a set of candidate regions with a high sensitivity ratio. The main focus of this work is the classification of the elements in the very unbalanced candidates set, by the use of Support Vector Machines (SVMs). We performed several experiments with different kernels and differently balanced training sets. The results obtained show that cost-sensitive SVMs trained with very unbalanced data sets achieve promising results in terms of sensitivity and specificity.

Key words: Lung Nodule Detection; Computer Aided Diagnosis; Multi-scale analysis; Support Vector Machines; Cost-sensitive classification.

1 Introduction and Materials

The chest radiography is by far the most common type of procedure for the initial detection and diagnosis of lung cancer; it is even preferred to more sensitive and precise techniques (e.g. MRI and CT) due to its non-invasivity characteristics, radiation dose and economic considerations. Several studies in the last two decades (e.g [1] and [2]) explain the difficulties of the technical production of the radiographic image and its correct diagnostic interpretation; this is also proved by impressive numbers reporting both diagnosticians' error rate and the patients' mortality. Since no improvement of these results has

Email addresses: campadelli@dsi.unimi.it (Paola Campadelli), casiraghi@dico.unimi.it (Elena Casiraghi), valentini@dsi.unimi.it (Giorgio Valentini).

been observed even when employing more sophisticated imaging techniques, the use of computer programs for radiographs analysis has been suggested by studies that also show the potentiality of early diagnosis improvement [3]. This is why in the last two decades a great deal of research work has been devoted to the development of systems aimed at lung nodules detection. Although a wide variety of them have been already proposed the problem is still open (see [5] for a review).
In this paper we describe the results obtained by our recently developed method which extracts a first set of candidate nodules and then classifies them to discard the false positives. The classification was performed using both Neural Networks (NN), with several architectures, and Support Vector Machines (SVMs), with different kernels and different settings of their parameters. Since true and false positives were greatly unbalanced, we applied a cost-sensitive approach to improve the sensitivity of the classifiers. We present only the results obtained with SVMs since they are the most robust and promising.

2 Candidate nodules extraction

We briefly sketch here the candidate extraction scheme which is composed of three consecutive processing steps.
At first the borders of the lung field are precisely defined by an algorithm which perform a multi-scale analysis of the image and works under no assumption. To get more information about the lung structure, a further processing is aimed at separating the *visible lung area* from those parts hidden behind the spine, the diaphragm and the heart (*not visible lung area*), where lung nodules may still be present (figure 1). The overall algorithm, and the comparison with other methods presented in the literature, are reported in [7] where it is shown that this is a good initialization step for a lung nodules detection system.
The second step is also based on a multi-scale framework; it processes the segmented lung area, to produce an *enhanced image* (figure 1) where the visibility of the nodules is much improved, notwithstanding their particular size, grey level and contrast characteristics. We think that a multi-scale approach is the missing part of the methods presented in the literature.
The image thus produced is the input of another method which extracts the nodule candidate regions at different radius values, and combines them. The result is a *regions image* (figure 1) whose pixel values are computed during the extraction scheme. Both the enhancement and the extraction schemes are described in details in [8]. The method has been tested on the standard $JSRT$ database [4], containing 247 radiographs: 93 of patients with no disease, and 154 containing lung nodules with different levels of detection subtleties and sizes. The overall number of regions extracted is 31100, about 125 regions per image, and only 5 true positives lost out of 154. The comparison of these

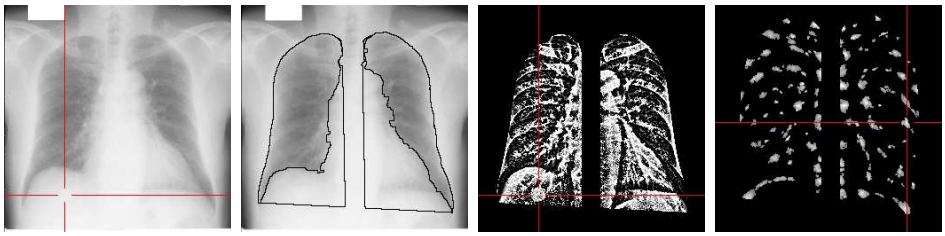


Fig. 1. Original Image containing a subtle nodule in the hidden areas; lung image area; enhanced image; regions image containing an extremely subtle nodule.

results with those of the only two methods presented in the literature, and tested on the same database ([10] and [9]) prove the better performance of the method [7].
To reduce the number of the extracted candidates we calculated for each region more than 100 features and performed a statistical analysis to select a set of 16 most representative ones (see [7] for their detailed description).
The novelty of this set is that it includes some features that are strictly related to the values computed during the extraction of the candidates; this is due to the observation of the strong dependency between the regions obtained and the algorithm used to extract them. Their efficacy is indeed proved by the fact that their combination by means of simple rules can discard more than 22000 candidates. The drawbacks of a rule based system are the empirically set thresholds used by the rules, which necessarily bring to a lack of robustness with respect to different databases. For this reason we experimented different learning machines such as NN and SVMs, using as input the same set of 16 features. In the following we will describe just the experiments with the SVMs which gave the most promising results.

3 Candidates classification using SVMs

In this section we present the results of the candidates classification performed with SVMs.
To evaluate the performance of a generic classifier two quantities are usually used. These are the sensitivity (sens.) and the specificity (spec.) and they are respectively defined as:

$$sens. = \frac{TP}{TP + FN} \qquad spec. = \frac{TN}{TN + FP} \tag{1}$$

where TP is the number of true positives, i.e. the positive examples correctly classified as positives, TN is the number of true negatives, FP is the number of false positives, i.e. the negative examples incorrectly classified as positives, and FN are the false negatives.
Our aim is to obtain a high sensitivity, in order to detect all the positive ex-

amples without a significant loss in specificity; from a medical point of view it is indeed crucial to detect all the positive examples, but at the same time we need to significantly reduce the number of false positives.

The data set is composed of 31100 nodule candidates (with only 149 positive examples), each candidate being represented by 16 features (see Sect. 2).
For the different experiments we performed a stratified random split of the data in training and test sets according to a train/test ratio equal to 3/2. We randomly repeated the above process 10 times, obtaining 10 pairs of training and test sets. In all the experiments we normalized the components of the data vectors to 0 mean and unitary standard deviation.
We applied SVMs with linear, polynomial and gaussian kernels, varying the regularization parameter (in the literature usually referred to as the C parameter) between 0.001 and 1000, the degree of polynomial kernels between 2 and 6 and the "width" (σ parameter) of gaussian kernels between 0.01 and 10000. We computed the mean error and standard deviation on the training and test sets, and the corresponding sensitivity and specificity. The train and test mean errors have been computed averaging between the 10 instances of the training and test set.
Considering the high unbalancing of the candidates set, at first we experimented with data sets "enriched" by positive examples, then we applied SVMs to more unbalanced data sets whose proportion, of positive to negative examples, is similar to the very unbalanced original data. Finally we applied a cost-sensitive approach to improve the sensitivity of SVMs.

3.1 Experiments with "positive enriched" data

For training and testing positive-enriched data sets were built, by considering separately positive and negative examples. We randomly split the available positive data in 89 examples for training and 60 examples for testing according to a train/test ratio equal to 3/2. From the set of negative data we extracted without replacement a number of negative examples equal to five times the number of positive data, both for the training and the test set, obtaining respectively $89 \times 5 = 445$ negative examples for the training set, and $60 \times 5 = 300$ negative examples for the test set.
The experiments generally gave poor results with all the models: even if the average test error generally obtained is quite good (on the average it's equal to 0.11) and the specificity is high (on the average it's equal to 0.96), the corresponding sensitivity is very low (between 0.41 and 0.48 in the models with the lowest test error). Even if we rank the SVM models with respect to the sensitivity level, in the best case we achieved a value of 0.53, losing about half of the true nodules (Table 1). Note that in Table 1 and in the following tables d stands for degree of polynomial kernels and σ refers to the width of

gaussian kernels.
We also ran experiments using a larger set of features, such as those generated through Gabor filters. Nevertheless, we obtained even worse results (data not shown).

SVM kernel	Sens.	Spec.	Test	Stdev	Train	stdev
Poly $d = 3$, $C = 200$	0.5333	0.9273	0.1383	0.0133	0.0579	0.0059
Poly $d = 2$, $C = 1000$	0.5317	0.9237	0.1417	0.0093	0.0607	0.0061
Poly $d = 4$, $C = 1000$	0.5317	0.8807	0.1775	0.0156	0.0139	0.0051
Gauss. $\sigma = 100$. $C = 1000$	0.5300	0.9003	0.1614	0.0195	0.0328	0.0080
Poly $d = 6$, $C = 200$	0.5300	0.8897	0.1703	0.0126	0.0150	0.0047
Poly $d = 5$, $C = 1000$	0.5300	0.8740	0.1833	0.0143	0.0047	0.0032

Table 1
Best SVM models with respect to the sensitivity ("positive enriched" data set).

3.2 Experiments with very unbalanced data

Analyzing the outputs of the SVMs, we observed that many positive examples are "strongly" classified as negative: that is, the corresponding real valued discriminant function is largely negative for many positive nodules. Even if, in principle, this may be due to the mislabeling of some examples, we suspected that a too small size of the training data might also play a significant role. To verify this hypothesis, we employed larger training sets, even if this necessarily involves a larger unbalance between positive and negative examples, since we have a very limited number of positive nodules.
We used a positive versus negative ratio equal to 1/30, hence obtaining respectively $89 \times 30 = 2670$ negative examples for the training set and $60 \times 30 = 1800$ negative examples for the test set. Even though, in this case, the SVMs can learn from more examples, the best sensitivity achieved is equal to 0.32, meaning that the training set is probably too unbalanced.

3.3 Experiments with cost-sensitive SVMs

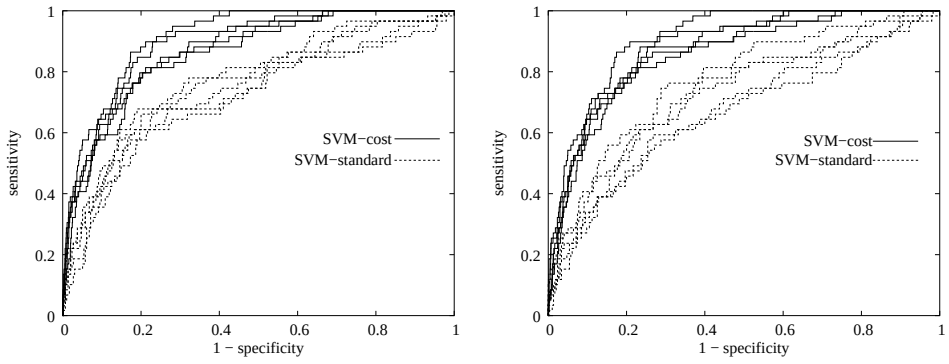
Based on the fact that we deal with very unbalanced data, we tried to outweigh errors on positive examples, introducing a cost-sensitive approach to improve the sensitivity of the SVMs.
In the framework of the SVM optimization problem we may introduce regularization parameters C_+ and C_- to be able to adjust the cost of misclassifications of false positives versus false negatives. Hence the minimization problem associated to linear 1-norm SVMs is translated into the following one, where asymmetrical loss functions are used [13]:

$$\begin{aligned} &\text{Minimize} && \mathbf{w}\cdot\mathbf{w} + C_+ \textstyle\sum_{i:y_i=+1} \xi_i + C_- \sum_{j:y_j=-1} \xi_j \\ &\text{subject to} && y_k(\mathbf{w}\cdot\mathbf{x}_k + b) \geq 1 - \xi_k \\ & && \xi_k \geq 0 \\ & && 1 \leq k \leq n \end{aligned} \tag{2}$$

where $(\mathbf{x}_k, y_k), \mathbf{x}_k \in \mathbb{R}^d, y_k \in \{-1, +1\}, 1 \leq k \leq n$, are the training examples, $\mathbf{w} \in \mathbb{R}^d$ is the weight vector, ξ_i and ξ_j are slack variables associated with positive and negative examples, and b is the constant factor of the linear function learned by the SVM.
In the experiments presented here we fixed $C_- = C$ and $C_+ = C \times C_f$, where C and C_f are respectively the regularization parameter and the cost-factor; we ran experiments where C_f was set equal to 2, 5, 10, 20, 50, 100, so that training errors on positive examples outweigh errors on negative examples.
Anyway, experiments with "positive-enriched" data sets achieved no significant increment in sensitivity at the expense of a significant decrement in specificity (data not shown).
On the contrary, applying a cost-sensitive approach with very unbalanced data (1/30 ratio between positive and negative examples) we achieved a significant increment in sensitivity with respect all the previous approaches (Table 2). Note that the larger test error obtained with cost-sensitive SVMs (Table 1 and 2) is mainly due to the unbalancing of the data. Anyway, our aim is to improve sensitivity in order to detect most of the real nodules: with relatively low values of C ($C < 0.01$) and quite large values of the cost factor C_f ($C_f \geq 50$), we obtained sensitivity equal or larger than 0.90 and specificity equal about to 0.70.
Figure 2 shows the Receiver Operating Characteristic (ROC) curves of cost-sensitive and standard polynomial and gaussian SVMs for five different splits of the training and test sets: cost-sensitive SVMs show better sensitivity and specificity compared with those of standard SVMs.

SVM kernel	Sens.	Spec.	Test	Stdev	Train	stdev
Poly $d = 7$, $C = 0.001$, $C_f = 100$	0.9458	0.5566	0.4311	0.0334	0.4215	0.0352
Poly $d = 3$, $C = 0.01$, $C_f = 100$	0.9458	0.5249	0.4618	0.0369	0.4481	0.0370
Poly $d = 4$, $C = 0.01$, $C_f = 100$	0.9288	0.5929	0.3964	0.0364	0.3864	0.0347
Gauss. $\sigma = 1000$. $C = 10$, $C_f = 100$	0.9254	0.5364	0.4512	0.0351	0.4423	0.0384
Linear $d = 6$, $C = 0.01$, $C_f = 100$	0.9254	0.5062	0.4805	0.0376	0.4750	0.0364
Poly $d = 4$, $C = 0.001$, $C_f = 50$	0.9051	0.6490	0.3429	0.0266	0.3339	0.0337
Gauss. $\sigma = 100$. $C = 2$, $C_f = 50$	0.8880	0.7202	0.2748	0.0201	0.2723	0.0162
Poly $d = 2$, $C = 0.1$, $C_f = 50$	0.8814	0.7007	0.2936	0.0174	0.2862	0.0158

Table 2
Results with very unbalanced data and asymmetrical cost functions: best SVM models with respect to the sensitivity.



(a) (b)

Fig. 2. Comparison of ROC curves in standard and cost-sensitive SVMs. (a) Polynomial SVMs (b) Gaussian SVMs.

4 Conclusions

The results show the effectiveness of cost-sensitive SVMs when dealing with highly unbalanced data sets, such as the one detected by a system for candidate nodules extraction. Anyway, this result is probably not sufficient for a population-wide clinical pre-screening of chest radiographs. Indeed the specificity obtained with high sensitivity is too low (about *0.70*) to be practical for diagnostic purposes. To increase the performances of our method, we plan to compute larger sets of features and employ feature selection techniques together with cost-sensitive SVM methods.

Acknowledgments

We would like to thank the anonymous reviewers for their precious comments and suggestions. This work has been developed in the context of the *CIMAINA* Center of Excellence.

References

[1] C. Vyborny, The aapm/rsna physics tutorial for residents: Image quality and the clinical radiographic examination, Radiographics 17 (2) (March 1997) 479–498.

[2] F. P. Stitik, Screening for cancer: Chest Radiology, Miller AB, edition, 1985.

[3] T.Matsumoto et al., Potential usefulness of computerized nodule detection in screening programs for lung cancer, Investigative Radiolology 27 (6) (June 1992) 471–475.

[4] J. Shiraishi, S. Katsuragawa, J. Ikezoe, T. Matsumoto, T. Kobayashi, K. Komatsu, M. Matsui, H. Fujita, Y. Kodera, and K. Doi , Development of a digital image database for chest radiographs with and without a lung nodule: receiver operating characteristic analysis of radiologists' detection of pulmonary nodules, American Journal of Roentgenology, 174 (1) (January 2000) 71–74.

[5] B. van Ginneken, B. ter H. Romeny, M. Viergever, Computer-aided diagnosis in chest radiography: A survey, IEEE Transaction on Medical Imaging, 20 (12) (December 2001) 1228–1241.

[6] H. Yoshida, X. Xu, K. Doi, M. Giger, Computer-aided diagnosis (cad) scheme for detecting pulmonary nodules using wavelet transforms, Proceedings of the International Society for Optical Engineering (SPIE) Medical Imaging: Image Processing, 2434 (May 1995) 621–626, Murray H. Loew Ed.

[7] E. Casiraghi, A Computer Aided Diagnosis System for Lung Nodules Detection in Postero Anterior Chest Radiographs, Phd. Dissertation, Università degli Studi, Milano, Italy, 7 March 2005.

[8] P. Campadelli, E. Casiraghi, Nodule detection in postero-anterior chest radiographs, Proceedings of Medical Image Computing and Computer Assisted Intervention, (MICCAI2004), Saint-Malo (France), September 27-31 2004, 3216 (II) 1048–1049, Springer.

[9] B. G. A. Schilham, M. Loog, Multi-scale nodule detection in chest radiographs, Proceedings of Medical Image Computing and Computer Assisted Intervention, (MICCAI2003), held in Montreal (Canada), November 15-18 2003, 2878 (I) 602–609, Springer.

[10] K. Bilgin, Y. Hiroyuki, Computerized detection of pulmonary nodules in chest radiographs based on morphological features and wavelet snake model, Medical Image Analisys, 6 (4) (December 2002) 431–447.

[11] B. van Ginneken, B. ter H. Romeny, Automatic segmentation of lung fields in chest radiographs, Medical Physics, 27 (10) (October 2000) 2445–2455.

[12] P. Campadelli, E. Casiraghi, Pruning the nodule candidate set in postero anterior chest radiographs, Proceedings of "Workshop Italiano Reti Neurali" (WIRN2004), Perugia (Italy), September 14-15 2004, Kluwer (in press).

[13] K.Morik, P. Brockhausen, T. Joachims, Combining statistical learning with a knowledge-based approach - a case study in intensive care monitoring, Proceedings of 16th International Conference on Machine Learning (ICML), Bled (Slovenia), June 27-30 1999, 268–277, Morgan Kaufmann Publishers.