Available online at www.sciencedirect.com
SCIENCE DIRECT®
Neurocomputing 68 (2005) 281–288
NEUROCOMPUTING
www.elsevier.com/locate/neucom

Letters

Support vector machines for candidate nodules classification

Paola Campadelli, Elena Casiraghi, Giorgio Valentini*

DSI—Dipartimento di Scienze dell'Informazione, Università degli Studi di Milano, Italy

Received 10 March 2005; received in revised form 31 March 2005; accepted 31 March 2005
Available online 24 May 2005
Communicated by R.W. Newcomb

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.

Keywords: 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

*Corresponding author. Tel.: +39 2503 16225; fax: +39 2503 16373.
E-mail addresses: campadelli@dsi.unimi.it (P. Campadelli), casiraghi@dico.unimi.it (E. Casiraghi), valentini@dsi.unimi.it (G. Valentini).

0925-2312/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.neucom.2005.03.005

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. [9,12]) 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 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 [5]. 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 [11] 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 (Fig. 1). The overall algorithm, and the comparison with other methods presented in the literature, are reported in [4] where it is shown that this is a good initialization step for a lung nodules detection system.

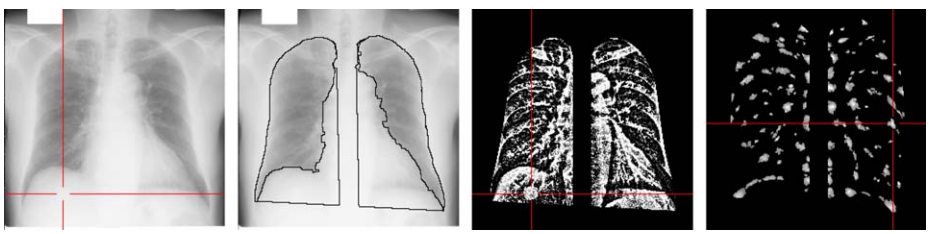

Fig. 1. Original image containing a subtle nodule in the hidden areas; lung image area; enhanced image; regions image containing an extremely subtle nodule.