A Multiclassifier Approach for Lung Nodule Classification

Carlos S. Pereira[1,2], Luís A. Alexandre[1,3],
Ana Maria Mendonça[1,4], and Aurélio Campilho[1,4]

[1] Instituto de Engenharia Biomédica
Rua Roberto Frias, 4200-465 Porto, Portugal
[2] Escola Superior de Tecnologia e Gestão de Lamego, Instituto Politécnico de Viseu
Av. Visconde Guedes Teixeira, 5100-074 Lamego, Portugal
cpereira@estgl.ipv.pt
[3] IT - Networks and Multimedia Group
6201-001 Covilhã, Portugal
lfbaa@di.ubi.pt
[4] Faculdade de Engenharia da Universidade do Porto
Rua Roberto Frias, 4200-465 Porto, Portugal
{amendon, campilho}@fe.up.pt

Abstract. The aim of this paper is to examine a multiclassifier approach to the classification of the lung nodules in X-ray chest radiographs. The approach investigated here is based on an image region-based classification whose output is the information of the presence or absence of a nodule in an image region. The classification was made, essentially, in two steps: firstly, a set of rotation invariant features was extracted from the responses of a multi-scale and multi-orientation filter bank; secondly, different classifiers (multi-layer perceptrons) are designed using different features sets and trained in different data. These classifiers are further combined in order to improve the classification performance. The obtained results are promising and can be used for reducing the false-positives nodules detected in a computer-aided diagnosis system.

1 Introduction

In the United States, lung and bronchus cancer is the second most common cancer and it is the leading cause for the estimated number of deaths [1]. Survival from lung cancer is directly related to early detection of suspicious lesions and treatment and successful results can be increased to 50% if the tumor is detected at an early stage. Posterior-anterior chest radiographs have been used by radiologists to make their diagnosis for some diseases, as pulmonary nodules, during a long time. However, studies show that radiologists detect pulmonary nodules in radiographs in about only 70-80% of actually positive cases [2].

Many computerized schemes have been developed for detection of pulmonary nodules in chest radiography, one of most studied problems in X-ray computer analysis. Most of the proposed computer-aided diagnosis (CAD) systems adopt

A. Campilho and M. Kamel (Eds.): ICIAR 2006, LNCS 4142, pp. 612–623, 2006.

a pattern recognition approach, combining a feature extraction phase with a classification step. The performance of the classifier depends directly on the discriminatory power of the feature set. Other systems follow a two-step approach that includes, first the selection of an initial set of lung nodule candidates and, second, reduction of false-positive candidates.

There is a vast amount of literature related with this subject. In 2001, Ginneken et al. [3], in a survey paper, distinguish three main areas, including general processing, segmentation of anatomical structures, and image analysis. The selection and classification of lung nodules is included in this last group, with references to 36 papers. The research on this topic continues, and the endeavor to propose new solutions to this difficult problem continues. In the following we briefly review some of the recent approaches.

Wei et al. [4] evaluated 210 features to look for the optimum feature set on 247 chest X-ray images. This CAD system consists of four processing steps: 1) location of tumor candidates by using adaptive ring filter, 2) extraction of the boundaries of tumor candidates, 3) extraction of feature parameters, and 4) discrimination between normal and abnormal regions. The authors report an average number of false-positives per image of 5.4 for a true positive detection rate of 80%.

Keserci et al. [2] describes an approach for the detection of lung nodules based on a combination of morphological features with an edge-guided wavelet snake model. With this combination the authors are able to largely reduce the number of false positives.

Ginneken et al. [5] presented a method to detect abnormalities in chest radiographs, using texture features, as the moments of responses to a multiscale filter bank. The authors report good results on a database of chest radiographs with interstitial diseases.

Schilham et al. [6] follow a multi-step approach for nodule detection. After lung segmentation the pre-candidate nodules are detected using a multiscale blob detector followed by a classifier to provide an initial reduction of nodule candidates. These candidates are further analyzed and a probability to represent a nodule is assigned to each blob, by using a k nearest neighbor classifier. The methodology was tested in JRST database [7] giving promising results.

Yoshida et al. [8] developed a method oriented to the reduction of false positives, exploring the symmetry between the two lungs and assuming that a nodule candidate region in one lung would correspond to a normal region in the other. These two regions are matched and difference in structure is evaluated. With this approach the authors are able to reduce the number of false-positives.

Suzuki et al. [9] report a reduction of false-positives, by using the so-called multiple massive-training neural network. The reported reduction is from 4.5 false positives to 1.4 per image, while maintaining the sensitivity. The input images are submitted to a background-trend-correction technique followed by contrast normalization.

In this paper is proposed a system for the classification of lung regions as nodular and non-nodular regions. The feature vector includes the original pixel

intensity and rotation invariant features extracted from the responses of a multi-scale and multi-orientation filtering process.

The layout of the rest of the paper is as follows: in Section 2, a brief description of the methodology is presented; Section 3 is dedicated to the filter bank used for feature extraction and the description of the feature vector; the different topologies and combinations of the classifiers are covered in Section 4, while Section 5 presents the experimental results of the training/testing phases of the classifiers. Finally, the conclusions are presented in Section 6.

2 Methodology

The proposed Pattern Recognition (PR) system is based on the following steps: 1) filtering of image regions using a multi-scale filter bank, 2) extraction of rotation invariant features and selection of the best discriminatory features, and 3) using multiple classifiers based on different multiple-layer perceptrons (MLP) receiving each one a different feature and combining the information from these several sources using several combination strategies. Figure 1 shows the block diagram representation of the processing scheme delineated for lung nodule region classification.

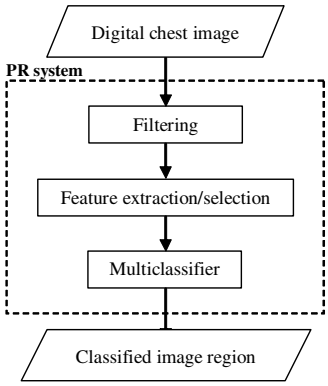

Fig. 1. Block diagram of the processing sequence for lung nodules region classification

3 Feature Measurement

Several approaches have been proposed for the segmentation and classification of texture contents in digital images. Chest radiographs are not an exception, as textural analysis has already been applied to the detection of pulmonary pathologies [3].

In this paper we propose a solution for the characterization of the texture of lung nodules based on a set of features derived from a multi-scale processing sequence that combines both spatial and frequency filters.

The basic assumption for most filtering approaches is that the energy distribution in the frequency domain identifies a texture. Hence, if the frequency spectrum of a textured image is decomposed into a sufficient number of subbands, the spectral energy signature of that texture is different from other distinct textures [10].

In [11], a bank of Gabor filters was used for extracting local image features. In the spatial domain, the Gabor function can be viewed as a sinusoidal plane of a particular frequency and orientation modulated by a Gaussian function. The Gabor function forms a complete but a non-orthogonal basis set and its impulse response in the 2-D plane has the following general form shown in equation (1), where u_0 denotes the radial frequency of the Gabor function [12].

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y} e^{\left[-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2}+\frac{x^2}{\sigma_x^2}\right)\right]} e^{2\pi j u_0 x} \tag{1}$$

The space constants σ_x and σ_y defines the Gaussian envelope along the x and y axes. A Gabor filter is circularly symmetric when $\sigma_x = \sigma_y$ and asymmetric when $\sigma_x \neq \sigma_y$. Each bank comprises a predefined number of Gabor filters that are the result of using different preferred spatial frequencies and different equidistant preferred orientations.

Another common approach for texture analysis is based on the processing of images with a large number of filters at multiple orientations and scales ("Gabor-like" filters) to extract features for classification purposes. However, constructing and storing a high dimensional filter response space is not computationally feasible and therefore it is necessary to limit the dimensionality of the filter response vector. This can be achieved if multiple oriented filters are used, but their output space is combined to form a low dimensional, rotation invariant response vector [13]. Leung and Malik filter set is a multi-scale, multi-orientation filter bank with 48 filters [14]. In this work, first and second derivatives of a Gaussian function are applied in 6 orientations and 3 scales making a total of 36 filters. The remaining filters comprise 8 Laplacian of Gaussian and 4 Gaussian kernels.

Another filter set is the Maximum Response 8 (MR8) filter bank which is composed by 38 filters. The MR8 bank contains an edge filter at 3 scales, and a bar filter at the same 3 scales; each of these basic kernels is rotated in order to generate 6 directional filters. The last 2 filters are a Gaussian filter and a Laplacian of Gaussian filter.

The main difference between the results of the MR8 bank and the Leung-Malik bank is rotational invariance, which is a unique characteristic of the former set. This attribute is obtained through the combination of the 6 directional responses associated with the 6 rotated versions the same basic filter into a single response, which is the maximum value across all directions. As a consequence, the responses of the 36 directional filters of the MR8 set are condensed into 6 single rotation invariant scale-dependent results.

Our proposal for lung nodule classification is also based on the use of a multi-scale filter bank, which is essentially an extension of the MR8 set, with the inclusion of two additional sets of scale-dependent Gaussian and Laplacian of

Gaussian kernels. This filter bank contains 48 filters that are a mix of edge, bar and spot kernels, at multiple scales and orientations, as shown in Fig. 2. The individual filters can be divided into two different classes: the first class comprises 6 sets of directional filters, shown in the left part of Fig. 2, whose individual results are combined to retain just the maximum absolute of the corresponding 6 directions; the second class consists of 12 isotropic filters, 6 Gaussian and 6 Laplacian of Gaussian, as depicted in the right part of the figure.

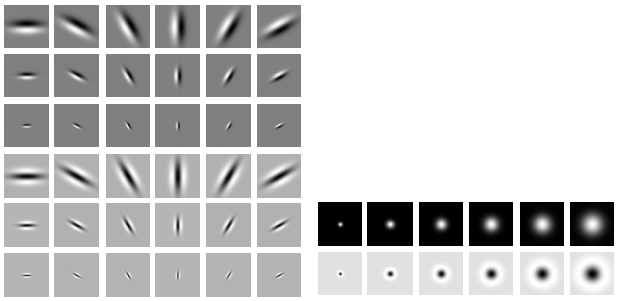

Fig. 2. Kernels of the proposed filter bank that comprises 36 filters at 6 different scales and orientations (left part of the figure) and 12 isotropic filters, 6 Gaussian and 6 Laplacian of Gaussian (right part of the figure)

To achieve rotational invariance, only the maximum filter response across all di-rections is kept. This means that although the filter bank is formed by 48 kernels, only 18 filters responses are finally used.

Each pixel of the image region under analysis is processed using the herein proposed filter bank, and, for each image point, the 18 filtering results are obtained. These results are combined with the pixel intensity, which is directly measured from the image, to generate a feature vector used for image region characterization. A graphical representation of the filtering sequence and feature measurement is depicted in Figure 3. Table 1 gives a brief description of the complete feature vector components.

4 Multiclassifier Approach

There are two main approaches to improve the performance of a pattern recognition system: the first is to consider features with high discriminatory power; the second is to improve the classifications result. The motivation of the use of a combination of classifiers is that classifiers with different methodologies, or that use different features, can complement each other. In this work, we have used multiple classifiers based on different multiple-layer perceptrons (MLP) receiving each one a different feature, and the information from these several sources is combined using arithmetic mean (or the sum), geometric mean (or the product), as in Alexandre et al. [15], minimum, maximum, borda, and median combination rules to produce the final classification decision.

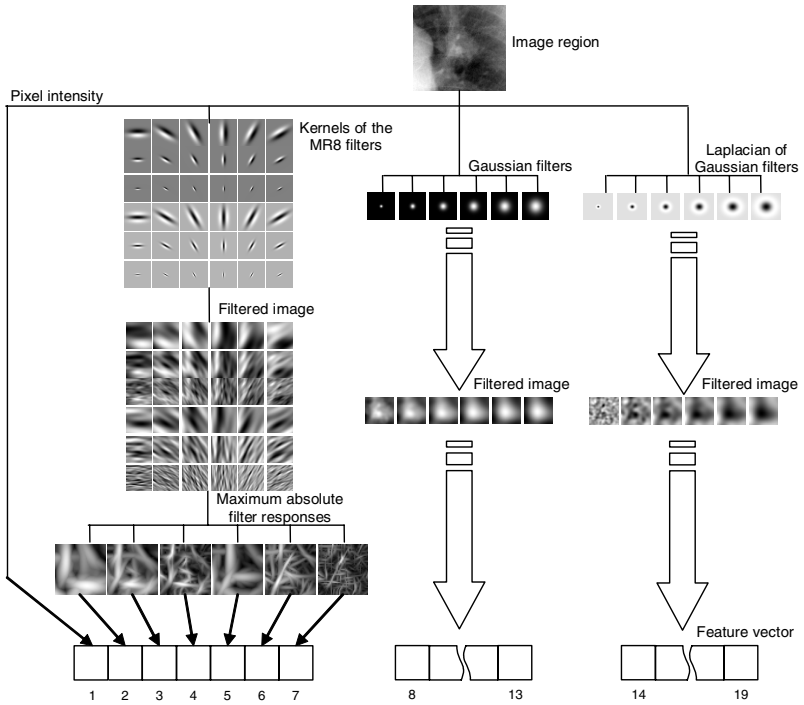

Fig. 3. Filtering sequence and feature measurement

Table 1. Description of the feature vector fields

Feature vector	
Index	**Description**
1	Pixel intensity at the original image
2-7	Edge and bar filter responses
8-13	Gaussian filter responses
14-19	Laplacian of Gaussian filter responses

The aim of the proposed image region-based nodule classification is producing the information of the presence or absence of a nodule for an entire region. For that, the following decisions are taken into account:

1. Each image region corresponds to a point in the data set;
2. Each data set corresponds to a MLP according to one of the features, with a topology selected after experiments;
3. Six combinations of classifiers are applied to the previous isolated MLP outputs to improve the performance of the final classification.

Since the input layer has a large number of pixels, corresponding to the size of the input image region, they are grouped into square regions of side X (each

region contains X^2 pixels). Every pixel in a group only has feed-forward connections to one pixel in the hidden layer. For instance, for 44×44 image regions and $X = 11$, the hidden layer will have its neurons organized in a square with size $44/X = 4$ elements, thus, the hidden layer will have 16 neurons.

Table 2 shows the number of weights in the neural network as a function of the number of the neurons in the hidden layer (NNHL) for two cases: net1 (the input layer fully connected to the hidden layer) and net2 (the input layer uses pixel groups where all the pixels in each of these groups only have a connection to one neuron in the hidden layer). The use of pixel groups in the input layer (net2) reduces, significantly, the number of weights used in the neural network, as shown in table 2.

Table 2. Number of weights in the neural network for different values of the number of neurons in the hidden layer, for the two types of network

NNHL	Net1	Net2
4	7757	1950
16	31025	1986
121	234620	2301

The results presented in table 3 shows that the network topology that yields better results is [1936:16:2]. Each topology was submitted to 40 repetitions.

Table 3. Error obtained using pixel intensity feature and different network topologies

NNHL	Error (std)
4	18.58 (2.49)
16	17.82 (2.16)
121	38.37 (16.75)

5 Experimental Results

Image regions of size 96×96 were automatically selected from chest radiographs of the JRST image database [7]. The size of the regions was established based on the maximum dimension of the nodules that are present in the database. To avoid the consideration of border conditions that naturally result from the filtering process, only the central 44×44 area (1936 pixels) was used for classification purposes. Three different sets for training/testing were formed: the first set contains regions centred in a true nodule area; the second set is composed by regions with no nodules that were randomly cropped out from pathologic images; finally, the third set holds re-gions with no nodules that were randomly extracted from images without any kind of pathology. Examples of the three image region sets are shown in Figure 4.

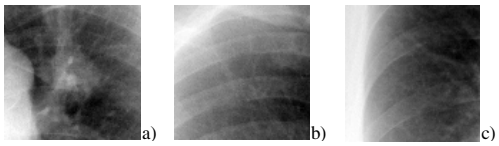

Fig. 4. Examples of image regions: a) region with a nodule in its central part; b) region without a nodule, extracted from a pathologic image; c) region without a nodule, extracted from a non-pathological image

In this section, the experimental results of image-based nodule detection are presented. As each image region will correspond to a point in the data set, a set of 19 MLPs were created, corresponding each data set to an image according to one of the features.

Table 4. Error, sensitivity and specificity for each classifier using a different feature as input, and their combinations

Classifier	**Error (std)**	**Sensitivity (std)**	**Specificity (std)**
1	16.21 (1.59)	76.41 (3.35)	88.39 (2.07)
2	23.07 (1.08)	66.23 (2.56)	83.60 (1.81)
3	22.73 (2.15)	70.35 (2.78)	81.58 (2.47)
4	23.98 (1.31)	67.75 (4.36)	81.17 (1.22)
5	31.21 (0.58)	60.06 (2.00)	74.22 (0.91)
6	33.92 (1.44)	59.96 (2.84)	69.91 (2.06)
7	38.61 (1.60)	51.08 (3.52)	67.81 (2.61)
8	18.12 (1.12)	78.03 (2.26)	84.28 (0.74)
9	17.37 (0.99)	79.44 (2.49)	84.62 (0.85)
10	19.08 (1.01)	75.43 (3.00)	84.35 (2.14)
11	20.32 (1.93)	73.48 (2.51)	83.54 (3.41)
12	21.20 (1.54)	70.02 (2.19)	84.28 (1.24)
13	22.65 (2.00)	65.69 (2.86)	84.62 (2.47)
14	30.63 (2.52)	61.47 (4.42)	74.29 (2.20)
15	25.60 (0.89)	70.35 (2.89)	76.92 (1.38)
16	20.37 (0.78)	76.19 (3.30)	81.78 (1.97)
17	19.87 (1.03)	74.35 (2.93)	83.74 (1.80)
18	19.62 (1.22)	75.54 (3.04)	83.40 (1.76)
19	19.87 (1.26)	74.35 (2.62)	83.74 (1.41)
Mean	16.54 (1.02)	77.49 (2.04)	87.18 (1.08)
Product	16.50 (0.81)	77.71 (2.08)	87.11 (0.79)
Minimum	17.91 (0.87)	75.76 (1.46)	86.03 (1.27)
Maximum	17.91 (0.88)	75.87 (1.39)	85.96 (1.30)
Borda	43.72 (1.49)	97.40 (1.09)	30.63 (2.56)
Median	16.54 (1.46)	77.49 (2.68)	87.18 (1.39)

Table 5. Confusion matrices for the experiments with 19 classifiers each one receiving a different feature

		Predicted values	
Classifier	**True values**	**Non-nodules**	**With nodules**
1	Non-nodules	218.33 (3.35)	28.67 (5.13)
	With nodules	36.33 (5.16)	117.67 (5.16)
2	Non-nodules	206.50 (4.46)	40.50 (4.46)
	With nodules	52.00 (3.95)	102.00 (3.95)
3	Non-nodules	201.50 (6.09)	45.50 (6.09)
	With nodules	45.67 (4.27)	108.33 (4.27)
4	Non-nodules	200.50 (3.02)	46.50 (3.02)
	With nodules	49.67 (6.71)	104.33 (6.71)
5	Non-nodules	183.33 (2.25)	63.67 (2.25)
	With nodules	61.50 (3.08)	92.50 (3.08)
6	Non-nodules	172.67 (5.09)	74.33 (5.09)
	With nodules	61.67 (4.37)	92.33 (4.37)
7	Non-nodules	167.50 (6.44)	79.50 (6.44)
	With nodules	75.33 (5.43)	78.67 (5.43)
8	Non-nodules	208.17 (1.83)	38.83 (1.83)
	With nodules	33.83 (3.49)	120.17 (3.49)
9	Non-nodules	209.00 (2.10)	38.00 (2.10)
	With nodules	31.67 (3.83)	122.33 (3.83)
10	Non-nodules	208.33 (5.28)	38.67 (5.28)
	With nodules	37.83 (4.62)	116.17 (4.62)
11	Non-nodules	206.33 (8.43)	40.67 (8.43)
	With nodules	40.83 (3.87)	113.17 (3.87)
12	Non-nodules	208.17 (3.06)	38.83 (3.06)
	With nodules	46.17 (3.37)	107.83 (3.37)
13	Non-nodules	209.00 (6.10)	38.00 (6.10)
	With nodules	52.83 (4.40)	101.17 (4.40)
14	Non-nodules	183.50 (5.43)	63.50 (5.43)
	With nodules	59.33 (6.80)	94.67 (6.80)
15	Non-nodules	190.00 (3.41)	57.00 (3.41)
	With nodules	45.67 (4.46)	108.33 (4.46)
16	Non-nodules	202.00 (4.86)	45.00 (4.86)
	With nodules	36.67 (5.09)	117.33 (5.09)
17	Non-nodules	206.83 (4.45)	40.17 (4.45)
	With nodules	39.50 (4.51)	114.50 (4.51)
18	Non-nodules	206.00 (4.34)	41.00 (4.34)
	With nodules	37.67 (4.68)	116.33 (4.68)
19	Non-nodules	206.83 (3.49)	40.17 (3.49)
	With nodules	39.50 (4.04)	114.50 (4.04)

Table 6. Confusion matrices for the experiments with the combinations of the 19 classifiers

		Predicted values	
Classifier	**True values**	**Non-nodules**	**With nodules**
Mean	Non-nodules	215.33 (2.66)	31.67 (2.66)
	With nodules	34.67 (3.14)	119.33 (3.14)
Product	Non-nodules	215.17 (1.94)	31.83 (1.94)
	With nodules	34.33 (3.20)	119.67 (3.20)
Minimum	Non-nodules	212.50 (3.15)	34.50 (3.15)
	With nodules	37.33 (2.25)	116.67 (2.25)
Maximum	Non-nodules	212.33 (3.20)	34.67 (3.20)
	With nodules	37.17 (2.14)	116.83 (2.14)
Borda	Non-nodules	75.67 (6.31)	171.33 (6.31)
	With nodules	4.00 (1.67)	150.00 (1.67)
Median	Non-nodules	215.33 (3.44)	31.67 (3.44)
	With nodules	34.67 (4.13)	119.33 (4.13)

The training and test procedure consists of 10 repetitions of the holdout: half of the set was used for training and the other half for testing. Then the sets were exchanged and the two errors on the test sets were averaged yielding one experiment result. The networks were trained using resilient backpropagation.

Table 4 shows the error, sensitivity and specificity for each one of 19 MLPs, with a [1936:16:2] topology, using a different feature as input, and their combinations. Tables 5 and 6 present the results of experiments using the previous 19 classifiers and their combinations, respectively. Each topology was submitted to 10 repetitions.

6 Conclusions

For the image region-based lung nodules classification, the following conclusions are drawn:

- The network topology that yields better results is [1936:16:2];
- The smallest error is always obtained using intensity as feature (classifier 1);
- The classifiers that present smallest error are, by decreasing error, those that use features 1, 9, 8 and with an almost identical error 10, 18, 17 and 19;
- The combinations cannot obtain an error smaller than that obtained with the best classifier (16.21%);
- All combinations, with the exception of the Borda, have better specificity than all isolated classifiers (with the exception of classifier 1);
- All combinations, with the exception of the Borda, show less errors than all classifiers, with the exception of classifiers 1 and 9;
- The Borda count combination method is the one that produces best values in terms of the sensitivity (97%), although with a high overall error (43%).

If we had to choose one of these approaches the best would be the combination using the product rule because although the error is higher than the best isolated classifier (16.50 versus 16.21) its standard deviation is significantly smaller (about one half: 0.81 versus 1.59). We would also like to point out one of the reasons we believe the multiclassifiers did not show a greater improvement over the isolated classifiers: we suspect that due to the large number of combined classifiers (19) the best possible fusion results were not achieved. We are currently considering the idea of choosing subsets of the 19 classifiers where we believe better results can be obtained.

As future work, the multi-classification approach proposed in this work will be used in combination with an initial step of candidate region selection, aiming at reducing the final number of false positives to be presented to the radiologist. The initial stage of candidate region selection is already implemented and some preliminary tests performed using the 154 nodular images of the JSRT database allowed the conclusion that the selection of 3 candidate regions per image identifies 46% of the nodules, while for 5 candidates this value is increased to 60%; an detection rate of 72% is achieved if 15 regions are initially selected.

References

1. A. Jemal, T. Murray, E. Ward, A. Samuels, R. C. Tiwari, A. Ghafoor, E. J. Feuer, and M. J. Thun, *Cancer Statistics, 2005*, CA Cancer J Clin, vol. 55, pp. 10-30, 2005
2. B. Keserci and H. Yoshida, *Computerized detection of pulmonary nodules in chest radiographs based on morphological features and wavelet snake model*, Medical Image Analysis, vol. 6, pp. 431-447, 2002
3. B. van Ginneken, Romeny, B.M.H., Viergever, M.A., *Computer-aided diagnosis in chest radiography: A survey*, IEEE Transactions on Medical Imaging, vol. 20, no. 12, 1228-1241, 2001
4. J. Wei, Y. Hagihara, A. Shimizu, and H. Kobatake, *Optimal image feature set for detecting lung nodules on chest X-ray images*, CARS 2002, 2002
5. B. van Ginneken, S. Katsuragawa, B. M. ter Haar Romeny, D. Kunio, and M. A. Viergever, *Automatic detection of abnormalities in chest radiographs using local texture analysis*, IEEE Transactions on Medical Imaging, vol. 21, pp. 139-149, 2002
6. A. M. R. Schilham, B. van Ginneken, and M. Loog, *Multi-scale Nodule Detection in Chest Radiographs*, Lecture Notes in Computer Science, vol. LNCS 2878, pp. 602-609, 2003
7. J. Shiraishi, S. Katsuragawa, J. Ikezoe, T. Matsumoto, T. Kobayashi, K.-i. 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, vol. 174, pp. 71-74, 2000
8. H. Yoshida, *Local contralateral subtraction based on bilateral symmetry of lung for reduction of false positives in computerized detection of pulmonary nodules*, IEEE Transactions on Biomedical Engineering, vol. 51, pp. 778-789, 2004
9. K. Suzuki, J. Shiraishi, H. Abe, H. MacMahon, and K. Doi, *False-positive reduction in computer-aided diagnostic scheme for detecting nodules in chest radiographs by means of massive training artificial neural network*, Academic Radiology, vol. 12, pp. 191-201, 2005

10. T. Randen, and J.H. Husoy, *Filtering for texture classification: A comparative study*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 21, no. 4, 1999
11. S.E. Grigorescu, N. Petkov, and P. Kruizinga, *Comparison of texture features based on Gabor filters*, IEEE Transactions on Image Processing, vol. 11, no. 10, pp. 195-203, 2002
12. A. Kumar, and G.K.H. Pang, *Defect detection in textured materials using Gabor filters*, IEEE Transactions on Industry Applications, vol. 38, no. 2, 2002
13. M. Varma, and A. Zisserman, *Unifying statistical texture classification frameworks*, Image and Vision Computing, vol. 22, pp. 1175-1183, 2004
14. T. Leung, and J. Malik, *Representing and recognizing the visual appearance of materials using three-dimensional textons*, International Journal of Computer Vision, vol. 43, pp. 29-44, 2001
15. Luís A. Alexandre, A.C. Campilho, and M. Kamel, *On combining classifiers using sum and product rules*, Pattern Recognition Letters, vol. 22, pp. 1283-1289, 2001