ESANN 2016 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 27-29 April 2016, i6doc.com publ., ISBN 978-287587027-8. Available from http://www.i6doc.com/en/.

Feature definition, analysis and selection for lung nodule classification in chest computerized tomography images

Luis Gonçalves [1,2], Jorge Novo [3], Aurélio Campilho[1,2]

1- INESC TEC - INESC Technology and Science
2- Faculdade de Engenharia, Universidade do Porto
FEUP Campus, Dr. Roberto Frias 4200 - 465. Porto (Portugal)
3- University of A Coruña, Department of Computer Science
Campus de Elviña, 15071. A Coruña (Spain)

Abstract.

This work presents the results of the characterization of lung nodules in chest Computerized Tomography for benign/malignant classification. A set of image features was used in the Computer-aided Diagnosis system to distinguish benign from malignant nodules and, therefore, diagnose lung cancer. A filter-based feature selection approach was used in order to define an optimal subset with higher accuracy.

A large and heterogeneous set of 293 features was defined, including shape, intensity and texture features. We used different KNN and SVM classifiers to evaluate the features subsets. The estimated results were tested in a dataset annotated by radiologists. Promising results were obtained with an area under the Receiver Operating Characteristic curve (AUC value) of $96.2 \pm 0.5\%$ using SVM.

1 Introduction

Nowadays, from all the pathologies, lung cancer is one of the most deadliest in the world, specially in developed countries. As a reference, the total amount of deaths of lung cancer in the United States is superior than the sum of colon, breast and prostate cancers [1]. A relatively poor early diagnosis is the main cause of dead, representing a crucial factor for the patient's survival rate. This happens because the diagnosis process is expensive and hard to obtain, as the radiologists must perform a deep and exhaustive revision throughout the scans, which implies a time consuming process and is often physically demanding [2][3].

The Computerized Tomography (CT) of the thorax is frequently used for the diagnosis of lung diseases due to a good image quality. Computer-aided Diagnosis (CAD) systems for lung nodules detection and characterization use a set of techniques, as for example progressive thresholds until obtaining the optimal nodule segmentation [4].

Once the lung nodule is segmented, an appropriate definition and characterization is crucial in order to analyze and determine its malignancy and, therefore, being able to proceed with lung cancer diagnosis. Thus, the features would include texture, gradient, morphology-based or intensity image information [4, 5, 6]. Then, feature selection can be useful to reduce the dimensionality

ESANN 2016 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 27-29 April 2016, i6doc.com publ., ISBN 978-287587027-8. Available from http://www.i6doc.com/en/.

of the feature space, avoiding redundant or irrelevant information. Finally, the classification stage aims to distinguish malignant and benign nodules. As an example, Yanjie et al. [6] employed genetic algorithms for feature selection and Support Vector Machines (SVM) based classifiers in the classification stage.

The systems can be analyzed to assess the performance using different statistics. One of the most used is the AUC value, which was employed in this work. The best results in the state-of-the-art for benign/malignant nodule classification were presented by Haifeng et al. [4] with an AUC value of 91% for diagnosed nodules by either biopsy, follow-up or surgery.

In this work, we used the diagnosis made by radiologist and a large set of heterogeneous features for lung nodule classification. A subset of features was selected in order to identify the most relevant discriminant ones. Moreover, different KNN and SVM classifiers were used to evaluate the potential of the feature subsets.

2 Methodology

Generally, the process of diagnosis includes a set of 5 main steps: nodule segmentation, feature measurement, feature selection, classification and validation. As input, the method used segmented nodules provided by an approach that was previously developed and described in [7].

2.1 Feature definition

In general terms, the main characteristics that the nodules present are related to intensity, shape and internal characteristics. These properties present different patterns that normally help the specialists to distinguish the benign from the malignant nodules.

Having all this in mind, we defined a large and heterogeneous set of 293 features, presented in Table 1, as input to the classifiers to discriminate the nodules regarding the malignancy. All these features can be categorized in the following groups:

Morphological-based features Properties as volume, compact, ratios between the eigenvalues or ratio of sphericity were defined to characterize appropriately the nodules margins, their shape or volume, [5][8]. Generally, benign nodules present a more compactness and smooth shape with respect to malignant ones, with higher irregular morphologies. These features are obtained from the 3D mask of the nodule.

Intensity features Different statistics over the entire intensity region as well as over concentric spheres with different radiuses were implemented. These properties basically measure the degree of calcification of the nodule where higher calcification implies higher intensity values, many times associated with benign nodules. The calcification is an important measure to differentiate benign from malignant nodules. These features are obtained from the segmented nodule after multiplying the 3D mask of the nodule with the original image.

ESANN 2016 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 27-29 April 2016, i6doc.com publ., ISBN 978-287587027-8.
Available from http://www.i6doc.com/en/.

Texture features Present high relevance as they indicate the internal characteristics of the nodules (air, fat, cavitation, etc.) helping, therefore, the benign/malignant differentiation. Gray-level intensity histogram (GLIH), the Gray-Level Co-Occurrence Matrix (GLCM) [4][5], Gabor filters and Laws' texture energy measures were included in the feature set. These features are obtained from the region of interest of the nodule.

Table 1: Measured features. (Gabor - the features are described as *Si-Ow*, where S is the scale, i is the total number of scales, O is the orientation and w is the total number of orientations. Other definitions: *max* - Maximum; *min* - Minimum; *std* - Standard Deviation; r - Radius)

	Features
Morphology based	[1] - Volume [2,3] - Compactness1; Compactness2 [4,5] - Ratio1; Ratio2 [6-9] - Eigen_ratio(1, 2, 3, 4) [10-12] - Sphericity_ratio1; Sphericity_ratio2; Sphericity_ratio3
Intensity	[13-18] - Overall Intensity: Max; Min; Mean; Median; Std [19-32] - Intensity Over Spheres (r = 1, 2, Eq_Sphere_Radius): Max; Min; Mean; Median; Std
Texture	*GLIH* [33-37] - Obliquity; Kurtosis; Energy; Entropy; Mean_intensity (μ)
	GLCM [33-53] - Autocorrelation; Contrast; Correlation; Cluster Prominence; Cluster Shade; Dissimilarity; Energy; Entropy; Homogeneity; Maximum probability; Sum of squares; Variance; Sum average; Sum variance; Sum entropy; Difference variance; Difference entropy; Information measure of correlation1; Information measure of correlation2; Inverse difference; Inverse difference normalized; Inverse difference moment normalized
	LAWS - Lattice Aperture Waveform Sets [58-68] - mean LAWS 3×3; std LAWS 3×3 (5 convolution masks) [59-85] - mean LAWS 5×5; std LAWS 5×5 (9 convolution masks)
	Gabor [86-166] - mean Gabor S5-O8; std Gabor S5 (5× 8 = 40 filters) [167-293] - mean Gabor S8-O8; std Gabor S8 (8× 8 = 64 filters)

2.2 Feature selection and classification

Once the feature set is defined, we proceed with the feature selection process in order to select those features that provided more information avoiding, therefore, redundancies and facilitating the classification process. Feature selection was performed by 10-fold cross validation using Correlation Feature Selection (CFS) algorithm, that analyses the strength of a feature in predicting the class of the object, but tends to give little importance to the inter-correlation of the features, and Relief-F algorithm, that samples instances randomly and checks the distance between them and the neighbours that have the same or different classes. A weight function uses the distances to rank them [9][10].

Regarding the classifiers, three SVM classifiers with exponential Kernels, where θ=[1,2,3] and three KNN with k=[13,15,17] were used. Classification was performed by 10-fold cross-validation with 50 repetitions, using the AUC as measure, and the mean and standard deviation of the AUC value were calculated.

ESANN 2016 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 27-29 April 2016, i6doc.com publ., ISBN 978-287587027-8. Available from http://www.i6doc.com/en/.

Table 2: Mean and standard deviation of 50 AUC % values, for 12 features selected by 2 model searches and 6 classifiers.

AUC (%)	13-KNN	15-KNN	17-KNN	1-SVM	2-SVM	3-SVM
CFS	93.2 ± 0.8	93.5 ± 0.9	94.1 ± 0.7	96.2 ± 0.5	96.3 ± 0.6	96.4 ± 0.5
Relief-F	94.7 ± 0.7	94.4 ± 0.8	94.4 ± 0.7	96.0 ± 0.6	96.3 ± 0.6	96.2 ± 0.6

3 Results and Discussion

To test the methods, different 3D scans were taken from the Lung Image Database Consortium and Image Database Resource Initiative (LIDC-IDRI) [11]. This public image database consists of diagnostic and lung cancer screening thoracic Computed Tomography scans with marked up, annotated lesions. These images contain a variable number of slices that compose the 3D scan, each slice with a resolution of 512×512. In this work, only solid nodules that were annotated by a consensus of 4 experienced radiologists are considered, resulting in a total of 179 malignant and 121 benign nodules.

Table 2 presents the values for each of the columns. The lowest value is achieved using 13-KNN (KNN with $k = 13$). Looking at the SVMs results with the CFS subset, there is no clear conclusion on which classifier is the best, though the 3-SVM (SVM with $\theta = 3$) is slightly better.

Feature selection indicated 12 features that obtained higher AUC values for all the classifiers. The majority of features chosen by both methods were texture features, though CFS also selected two shape features (Volume and Compactness) and Relief-F three intensity (maximum intensity from an equivalent sphere, centered in the nodule's position with radius equal to 1, 2 and radius of the nodule). The CFS selected four GLCM and four Laws features. The inclusion of Volume and Compactness by CFS is coherent, as radiologists tend to consider small and round nodules as benign and big or spiculated as malignant. The lack of intensity features can give two indications, one is that CFS finds intensity information similar for both malignant and benign nodules, and the other is that the intensity features can be simply redundant to the problem if there are already corresponding high correlated features in the set. Relief-F, however, selected features that focus on the center calcification of the nodules, which are also very used by radiologists in predicting the malignancy of the nodules. It also includes a lot of GLCM features (8 in 12 features), implying that the GLCM has a great discrimination capacity.

The AUC values for all classifiers and subsets for this dataset are high, though there is no considerable difference between them. The reason behind this may be due to the fact that radiologists tend to classify nodules considering only morphological characteristics, as benign and malignant nodules generally have specific shapes and sizes. For example, if the nodule is spiculated or big then radiologists tend to classify the nodule as malignant. If it is small and round then radiologists most likely think the nodule is benign.

Different examples of correct classification results are depicted in Figure 1. It also represents the confidence as the amount of certainty that a classifier has

ESANN 2016 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 27-29 April 2016, i6doc.com publ., ISBN 978-287587027-8.
Available from http://www.i6doc.com/en/.

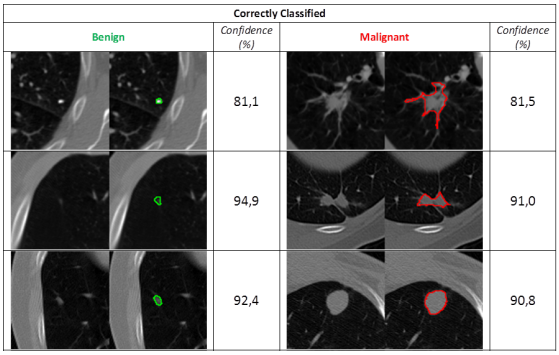

Fig. 1: Examples of correct benign (1^{st} column) and malignant (2^{nd} column) nodules classifications. Confidence is presented as the posterior probability of a nodule belonging to a particular class (benign or malignant).

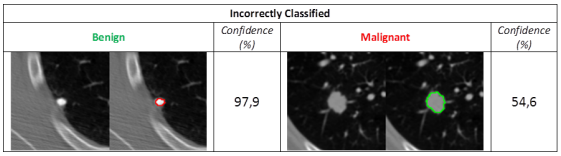

Fig. 2: Missclassification results of a benign (1^{st} column) and a malignant (2^{nd} column). Confidence is presented as the posterior probability of a nodule belonging to a particular class (benign or malignant).

on labeling a nodule as benign or malignant. Most of the nodules were correctly classified, as those in Figure 1. However, there were some fails that are illustrated in Figure 2, one benign nodule that was identified as malignant and one malignant as benign. The confidence of the classifier presents valuable and clear information of what happens in the system and visually confirms what was stated previously. It is visible that round, small nodules are labeled as benign with a high confidence and the same happens to big, spiculated/lobulated nodules. In Figure 2, the incorrect malignant classifications (on the right) happen in instances where one case has two or more nodules, possibly indicating a metastasis situation, leading radiologists to assume the nodules are malignant. The classifiers do not incorporate this information so they make a different diagnosis. The incorrect benign classifications happen in nodules that are large, spiculated or are located near or connected to other lung structures, leading to poor feature measurement.

4 Conclusions

This work tries to evaluate and select the best performance of the characteristics of pulmonary nodules in chest CT scans. As a result of this analysis, a large set of features were defined over the nodules in order to discriminate the benign from

ESANN 2016 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 27-29 April 2016, i6doc.com publ., ISBN 978-287587027-8. Available from http://www.i6doc.com/en/.

the malignant nodules. Moreover, this set was analyzed and feature selection was performed to select those ones that provided the best behaviour, by avoiding the redundancies in the feature definition and optimize the classification process. The posterior classification step achieved a performance with an AUC value of $96.2 \pm 0.5\%$, indicating that a CAD system can have similar classification performance as a radiologist. However, further analysis and additional validation is needed due to the small cases set that was employed. Future works will analyze and employ more sophisticated image features to provide the classification stage with more discriminative power. Moreover, wrapper based feature selection methods should be tested, as better results can be achieved by guiding the selection using classifiers.

Acknowledgements

This work is supported by the European ERDF and ESF funds, and Portuguese funds through the COMPETE and POPH programs of the Fundação para a Ciência e a Tecnologia (FCT), with the UID/EEA/50014/2013 project, and the grant contract SFRH/BPD/85663/2012.

References

[1] R.T. Greenlee, T. Murray, S. Bolden, P.A. Wingo. Cancer statistics. *CA: A Cancer Journal for Clinicians 2000*, 50:7–33, 2000.

[2] C. J. Breadsmoore and N. J. Screaton. Classification, staging and prognosis of lung cancer. *European Journal of Radiology*, 45:8–17, 2003.

[3] B. van Ginneken. *Computer-Aided Diagnosis in Thoracic Computed Tomography*, volume 12, pages 11–22. 2008.

[4] Haifeng Wu and Tao Sun et al. Combination of radiological and gray level co-occurrence matrix textural features used to distinguish solitary pulmonary nodules by computed tomography. *Society for Imaging Informatics in Medicine*, 26:797–802, 2013.

[5] Hui Chen and Yan Xu et al. Neural network ensemble-based computer-aided diagnosis for differentiation of lung nodules on ct images. *Academic Radiology*, 17(5):248–257, 2010.

[6] Yanjie Zhu and Yongqiang Tan et al. Feature selection and performance evaluation of support vector machine (svm)-based classifier for differentiating benign and malignant pulmonary nodules by computed tomography. *J. of Digital Imaging*, 23(1):51–65, 2010.

[7] Jorge Novo, Luís Gonçalves, Ana Maria Mendonça, and Aurélio Campilho. 3d lung nodule candidates detection in multiple scales. In *MVA 2015-IAPR International Conference on Machine Vision Applications*, pages 5–8, 2015.

[8] K. Murphy and B. van Ginneken et al. A large-scale evaluation of automatic pulmonary nodule detection in chest ct using local image features and k-nearest-neighbour classification. *Elsevier-Medical Image Analysis*, 13:757–770, 2009.

[9] Igor Kononenko. Estimating attributes: analysis and extensions of relief. In *Machine Learning: ECML-94*, pages 171–182. Springer, 1994.

[10] Mark A Hall. *Correlation-based feature selection for machine learning*. PhD thesis, The University of Waikato, 1999.

[11] S.G. Armato and G. McLennan et al. The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): A completed reference database of lung nodules on CT scans. *Medical Physics*, 38:915–931, 2011.