Comparison of FLDA, MLP and SVM in Diagnosis of Lung Nodule

Aristófanes Corrêa Silva[1], Anselmo Cardoso de Paiva[2], and Alexandre Cesar Muniz de Oliveira[2]

[1] Federal University of Maranhão - UFMA, Department of Electrical Engineering
Av. dos Portugueses, SN, Campus do Bacanga, Bacanga
65085-580, São Luís, MA, Brazil
ari@dee.ufma.br
[2] Federal University of Maranhão - UFMA, Department of Computer Science
Av. dos Portugueses, SN, Campus do Bacanga, Bacanga
65085-580, São Luís, MA, Brazil
paiva@deinf.ufma.br, acmo@deinf.ufma.br

Abstract. The purpose of the present work is to compare three classifiers: Fisher's Linear Discriminant Analysis, Multilayer Perceptron and Support Vector Machine to diagnosis of lung nodule. These algorithms are tested on a database with 36 nodules, being 29 benigns and 7 malignants. Results show that the three algorithms had similar performance on this particular task.

Key words: Fisher's Linear Discriminant Analysis, Multilayer Perceptron, Support Vector Machine, Diagnosis of Lung Nodule

1 Introduction

Lung cancer is known as one of the cancers with shortest survival after diagnosis [1]. Therefore, the sooner it is detected the larger the patient's chance of cure. On the other hand, the more information physicians have available, the more precise the diagnosis will be.

Solitary lung nodules are an approximately round lesion less than 3 cm in diameter and completely surrounded by pulmonary parenchyma. Larger lesions should be referred to as pulmonary masses and should be managed as likely malignant. In this situation, prompt diagnosis and staging are necessary to choose the proper treatment [1].

On the other hand, manage a pulmonary nodule is always a challenge, because in spite of benign possibility becoming more important as the nodule's dimension decreases, malignity has always to be excluded. If malignity has more than 5% of chance to be present a biopsy method must be indicated and for more than 60% the pacient is sent directly to ressection. Less than 5% allows a close following to prove stability. A recent problem, which has been becoming more frequent nowadays, is that Computerized Tomography

(CT) is finding nodules not visibles in a conventional chest X-ray in high risk groups (ie. heavy smokers) and frequently their little dimensions make difficult or impossible a biopsy procedure. On the other side, systematic ressection would increase unnecessary surgery at unacceptables levels. In this circumstance, a quantitative image method of volumetric determination are becoming recognized as an important parameter to stablish following criteria.

Macroscopicaly, lung nodules have a very variable tissue's structure. There can be nodules with tissue alterations almost imperceptible to the human eye and others presenting very noticeable alterations. The diagnosis gold standard is the histological examination, but image methods and in special Computed X-ray Tomography can aid diagnostic process in analyzing nodule's atributes like shape, presence and pattern of calcifications, walls of cavitations, aerobronchogram and, more recentely, mean attenuation coefficient before and after intravenous contrast standardized injection.

However, besides numerous reports of qualitative morphologic CT data in medical literature, there are relatively few reports of quantitative CT data and it seems that, in general, they are underutilized. We hipothetized that quantitative CT data derived from geometric and texture parameters may contribute to differential diagnosis between benign and malignant solitary pulmonary nodule, even without contrast utilization. The top row in Figure 1 shows the texture for two benign (a and b) and two malignant (c and d) nodules. The bottom row in Figure 1 shows the shape for two benign (a and b) and two malignant (c and d).

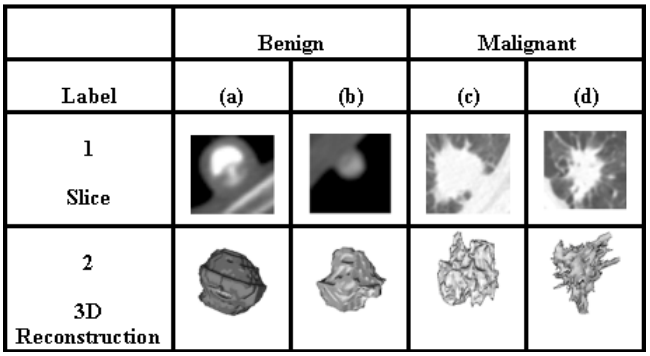

Fig. 1. Examples of benign lung nodules and malignant lung nodules.

The purpose of the present work is to compare three classifiers: Fisher's Linear Discriminant Analysis, Multilayer Perceptron and Support Vector Machine to diagnosis of lung nodule. Features extracted of nodules are based on CT images and analysis is supplied regarding the 3D geometry of the nodule. The validation of the classifiers is done by means of leave-one-out technique.

The analysis and evaluation of tests was done using the area under the ROC (Receiver Operation Characteristic) [2] curve.

2 Methods

2.1 Image Acquisition

The images were acquired with a Helical GE Pro Speed tomography under the following conditions: tube voltage 120 kVp, tube current 100 mA, image size 512×512 pixels, voxel size $0.67 \times 0.67 \times 1.0$ mm. The images were quantized in 12 bits and stored in the DICOM format [3].

2.2 3D Extraction and Reconstruction of Lung Nodules

In most cases, lung nodules are easy to be visually detected by physicians, since their shape and location are different from other lung structures. However, the nodule's voxel density is similar to that of other structures, such as blood vessels, which makes automatic computer detection difficult. This happens especially when a nodule is adjacent to the pleura. For these reasons, we have used the 3D region-growing algorithm with voxel aggregation [4]to make the nodule detection, which provides physician greater interactivity and control over the segmentation and determination of required parameters (thresholds, initial and final slice, and seed).

Two other resources provide greater control in the segmentation procedure: the barrier and the eraser. The barrier is a cylinder placed around the nodule by the user with the purpose of restricting the region of interest and stopping the segmentation by voxel aggregation from invading other lung structures. The eraser is a resource of the system that allows physicians to erase undesired structures, either before or after segmentation, in order to avoid and correct segmentation errors [5]. The bottom row in Figure 1 shows the 3D reconstruction of the nodules in the top row and exemplifies the nodule segmentation.

2.3 Lung Nodule Features

Skeletonization is a convenient tool to obtain a simplified representation of shapes that preserves most topological information [6]. A skeleton captures the local symmetry axes and is therefore centered in the object. In image analysis, features extracted from skeletons are commonly used in pattern-recognition algorithms [7]. Skeletons contain information about shape features which are very important in this work context.

We have used Zhou and Toga's algorithm [8] in the skeletonization process. They have proposed a voxel-coding approach to efficiently skeletonize volumetric objects. Each object point has two codes. One is the Boundary Seeded code (BS), which coincides with the traditional distance transform to indicate the minimum distance to the object's boundary. The second code is the so-called Single Seeded

code (SS), which indicates the distance to a specific reference point. SS code is used to extract the shortest path between a point in the object and the reference point. These paths are represented by sequential sets of voxels that will compose the initial skeleton. The key idea of voxel coding is to use the SS codes to generate a connected raw skeleton and the BS codes to assure the centeredness of the final skeleton.

Figure 2 shows the application the skeleton algorithm based on the nodules in Figure 1(a), (b), (c) and (d), respectively. It is easy observe that malignant nodules have more segments than benign nodules.

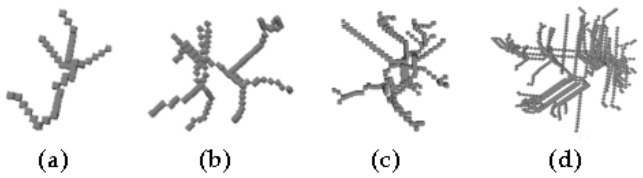

Fig. 2. Application of the skeleton algorithm based on the nodules in Figure 1(a), (b), (c) and (d).

We have extracted ten measures based on skeletons to analyze lung nodules, six of them have been used to describe the nodule's geometry. They are :

a) Number of Segments (NS)
b) Number of Branches (NB)
c) Fraction of Volume (FV): FV is defined by

$$FV = \frac{v}{V}$$

where v is the skeleton volume and V is the lung nodule's volume.
d) Length of Segments (LS):Defined by

$$LS = \frac{L}{\sqrt[3]{V}}$$

where L is the length of all segments and V is the lung nodule's volume.
e) Volume of the Skeleton Convex Hull (VCH)
f) Rate between the number of segments and the volume of the convex hull (NSVCH) [7]
Trying to charcaterize the nodule based on the skeleton texture we compute the density histogram of the N larger skeleton segments, where N is the smaller number of segments in the nodule's database. From this histogram e compute:
g) Variation Coefficient (VC): The VC is a measure of relative dispersion and is given by

$$VC = \frac{\sigma}{\mu}$$

, where σ is the standard deviation and μ is the mean.

h) Histogram Moments (variance (M_2), skewness (M_3), kurtosis (M_4)) defined as:

$$M_n = \frac{\sum (x_i - \mu)^n f_i}{N} \tag{1}$$

where $n = 2, 3, 4$, μ is the mean, N denotes the number of voxels in the segment, and f_i is the histogram.

More detailed information on moment theory can be found in [9].

2.4 Classification Algorithms

A wide variety of approaches has been taken towards the classification task. Three main historical strands of research can be identified [10]: statistical, neural network and machine learning. This section give an overview of Fisher's Linear Discriminant Analysis, Multilayer Perceptron and Support Vector Machine based on paradigms cited above.

Fisher's Linear Discriminant Analysis - FLDA: Linear discrimination, as the name suggests, looks for linear combinations of the input variables that can provide an adequate separation for the given classes. Rather than look for a particular parametric form of distribution, LDA uses an empirical approach to define linear decision planes in the attribute space i.e. it models a surface. The discriminant functions used by LDA are built up as a linear combination of the variables that seek to somehow maximize the differences between the classes [11]:

$$y = \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n = \beta^{'} x \tag{2}$$

The problem then reduces to find a suitable vector β. There are several popular variations of this idea, one of the most successful being the Fisher Linear Discriminant Rule. Fisher's Rule is considered a "sensible" classification, in the sense that it is intuitively appealing. It makes use of the fact that distributions that have a greater variance between their classes than within each class should be easier to separate. Therefore, it searches for a linear function in the attribute space that maximizes the ratio of the between-group sum-of-squares (B) to the within-group sum-of-squares (W). This can be achieved by maximizing the ratio

$$\frac{\beta' B \beta}{\beta' W \beta} \tag{3}$$

and it turns out that the vector that maximizes this ratio, β, is the eigenvector corresponding to the largest eigenvalue of $W^{-1}B$ i.e. the linear discriminant function y is equivalent to the first canonical variate. Hence the discriminant rule can be written as:

$$x \in i \text{ if } \left|\beta^T x - \beta^T u_i\right| < \left|\beta^T x - \beta^T u_j\right|, \text{for all } j \neq i \tag{4}$$

where $W = \sum n_i S_i$ and $B = \sum n_i (x_i - x)(x_i - x)^{'}$, and n_i is class i sample size, S_i is class i covariance matrix, x_i is the class i mean sample value and x is the population mean.

Stepwise discriminant analysis [11] was used to select the best variables to differentiate between groups. These measures were used in the FLDA, MLP and SVM classifiers.

Multilayer Perceptron: The Multilayer Perceptron - MLP, a feed-forward back-propagation network, is the most popular neural network technique in pattern recognition [12], [13]. Briefly, MLPs are supervised learning classifiers that consist of an input layer, an output layer, and one or more hidden layers that extract useful information during learning and assign modifiable weighting coefficients to components of the input layers. In the first (forward) pass, weights are assigned to the input units and to the nodes in the hidden layers and between the nodes in the hidden layer and the output, determine the output. The output is compared with the target output. An error signal is back propagated and the connection weights are adjusted correspondingly. During training, MLPs construct a multidimensional space, defined by the activation of the hidden nodes, so that the two classes (benign and malignant nodules) are as separable as possible. The separating surface adapts to the data.

Support Vector Machine: The Support Vector Machine (SVM) introduced by V. Vapnik in 1995 is a method to estimate the function classifying the data into two classes [14], [15]. The basic idea of SVM is to construct a hyperplane as the decision surface in such a way that the margin of separation between positive and negative examples is maximized. The SVM term comes from the fact that the points in the training set which are closest to the decision surface are called support vectors. SVM achieves this by the structural risk minimization principle that is based on the fact that the error rate of a learning machine on the test data is bounded by the sum of the training-error rate and a term that depends on the Vapnik-Chervonenkis (VC) dimension.

The process starts with a training set of points $x_i \in \Re^n, i = 1, 2, \cdots, l$ where each point x_i belongs to one of two classes identified by the label $y_i \in \{-1, 1\}$. The goal of maximum margin classification is to separate the two classes by a hyperplane such that the distance to the support vectors is maximized. The construction can be thought as follow: each point x in the input space is mapped to a point $z = \Phi(x)$ of a higher dimensional space, called the feature space, where the data are linearly separated by a hyperplane. The nature of data determines how the method proceeds. There are data that are linearly separable, nonlinearly separable and with impossible separation. This last case be still tracted by the SVM. The key property in this construction is that we can write our decision function using a kernel function $K(x, y)$ which is given by the function $\Phi(x)$ that maps the input space into the feature space. Such decision surface has the

equation:

$$f(x) = \sum_{i=1}^{l} \alpha_i y_i K(x, x_i) + b \tag{5}$$

where $K(x, x_i) = \Phi(x).\Phi(x_i)$, and the coefficients α_i and the b are the solutions of a convex quadratic programming problem [14], namely

$$\begin{array}{rl} \min\limits_{w,b,\xi} & \frac{1}{2} w^T \cdot w + C \sum\limits_{i=1}^{l} \xi_i \\ \text{subject to} & y_i \left[w^T \cdot \phi\left(x_i\right) + b \right] \geq 1 - \xi_i \\ & \xi_i \geq 0. \end{array} \tag{6}$$

where $C > 0$ is a parameter to be chosen by the user, which corresponds to the strength of the penalty errors, and the ξ_i's are slack variables that penalize training errors.

Classification of a new data point x is performed by computing the sign of the right side of Equation 5. An important family of kernel functions is the Radial Basis Function, more commonly used for pattern recognition problems [14], which has been used in this paper and is defined by:

$$K(x, y) = e^{-\gamma \|x-y\|^2} \tag{7}$$

where $\gamma > 0$ is a parameter that also is defined by the user.

2.5 Validation and Evaluation of the Classification Methods

In order to validate the classificatory power of the discriminant function, the leave-one-out technique [16] was employed. Through this technique, the candidate nodules from 35 cases in our database were used to train the classifier; the trained classifier was then applied to the candidate nodules in the remaining case. This technique was repeated until all 36 cases in our database had been the "remaining" case.

In order to evaluate the ability of the classifier to differentiate benign from malignant nodules, the area (AUC) under the ROC (Receiver Operation Characteristic) [2] curve was used. In other words, the ROC curve describes the ability of the classifiers to correctly differentiate the set of lung nodule candidates into two classes, based on the true-positive fraction (sensitivity) and false-positive fraction (1-specificity). Sensitivity is defined by $TP/(TP + FN)$, specificity is defined by $TN/(TN + FP)$, and accuracy is defined by $(TP + TN)/(TP + TN + FP + FN)$, where TN is true-negative, FN is false-negative, FP is false-positive, and TP is true-positive.

3 Results

The tests described in this paper were carried out using a sample of 36 nodules, 29 benign and 7 malignant. It is important to note that the nodules were diagnosed

by physicians and that the diagnosis was confirmed by means of surgery or based on their evolution. Such process takes about two years, which explains the reduced size of our sample.

There were no specific criteria to select the nodules. The sample included nodules with varied sizes and shapes, with homogeneous and heterogeneous characteristics, and in initial and advanced stages of development.

SPSS (*Statistical Package for the Social Sciences*) [17], LIBSVM [18] and NeuralPower [19] were used to training and classification of lung nodules to FLDA, MLP and SVM, respectively. ROCKIT [20] software was used to compute and compare the area under the ROC curve.

Stepwise discriminant analysis [11] was used to select the best variables to differentiate between groups, and the measures selected were NS, VCH and VC (with N as 1). These measures were used to FLDA, MLP and SVM classifiers.

We use the following parameters in the MLP classifier: one hidden layer with four units, hiperbolic tangent as the activation function, the value of 0.15 for the learning ratio, the value of 0.75 for the momentum.

In the classification via SVM a proposed procedure by the authors of LIBSVM [18] was used to obtain the best constants C and γ with a process of 36-fold cross-validation. In our case, $C = 2.0$ and $\gamma = 2.0$.

Table 1 shows the results of studied classifiers applied to nodule's 3D geometry. Based on the area of the ROC curve, we have observed that: (i) All classifiers have value AUC above 0.800, which means results with accuracy between good and excellent [21]. (ii) SVM have the minor value of sensitivity. (iii) The difference between the ROC curve using the FLDA and the MLP classifiers did not reach statistical significance ($p = 0.641$). The difference between the ROC curve using the FLDA and the SVM classifiers did not reach statistical significance ($p = 0.523$). The difference between the ROC curve using the MLP and the SVM classifiers did not reach statistical significance ($p = 0.799$).

Classifiers	Specificity %	Sensitivity %	Accuracy %	$AUC \pm$ SE
FLDA	89.7	71.4	86.1	0.946 ± 0.061
MLP	89.7	85.7	88.8	0.906 ± 0.079
SVM	89.7	57.1	83.3	0.892 ± 0.084

Table 1. Analysis of FLDA, MLP and SVM classifiers.

The number of nodules studied in our dataset is too small to allow us to reach definitive conclusions, but preliminary results from this work are very encouraging, demonstrating the potential for multiple variables used in a pattern classification approach to discriminate benign from malignant lung nodules.

4 Conclusion

FLDA, MLP and SVM have been applied to many classifications problems, generally yielding good performance. In this paper, we have compared these three classification algorithms on diagnosis of lung nodule. Results based on the analysis of the ROC curve have shown that the three algorithms had similar performance on this particular task. But a more accurate analysis of the SVM shows that it results in a not so good sensitivity, being less apropriated for a clinical use. Based on these results, we have observed that such measures provide significant support to a more detailed clinical investigation, and the results were very encouraging when nodules were classified with these classifiers. Nevertheless, there is the need to perform tests with a larger database and more complex cases in order to obtain a more precise behavior pattern.

Despite the good results obtained only by analyzing the geometry, further information can be obtained by analyzing the texture. As a future work, we propose a combination of texture and geometry measures for a more precise and reliable diagnosis.

Acknowledgments

We would like to thank Dr. Rodolfo A. Nunes and his team for the clinical support, and the staff from Instituto Fernandes Figueira, particularly Dr. Marcia Cristina Bastos Boechat, for the images provided.

References

1. Tarantino, A.B.: 38. In: Nódulo Solitário Do Pulmão. 4 edn. Guanabara Koogan, Rio de Janeiro (1997) 733–753
2. Erkel, A.R.V., Pattynama, P.M.T.: Receiver operating characteristic (ROC) analysis: Basic principles and applicattions in radiology. European Journal of Radiology **27** (1998) 88–94
3. Clunie, D.A.: DICOM Structered Reporting. PixelMed Publishing, Pennsylvania (2000)
4. Nikolaidis, N., Pitas, I.: 3-D Image Processing Algorithms. John Wiley, New York (2001)
5. Silva, A.C., Carvalho, P.C.P.: Sistema de análise de nódulo pulmonar. In: II Workshop de Informática aplicada a Saúde, Itajai, Universidade de Itajai (2002) Available at http://www.cbcomp.univali.br/pdf/2002/wsp035.pdf.
6. Gonzalez, R.C., Woods, R.E.: Digital Image Processing. 3 edn. Addison-Wesley, Reading, MA, USA (1992)
7. da F. Costa, L., Velte, T.J.: Automatic characterization and classification of glangion cells from the salamander retina. The Journal of Comparative Neurology **404** (1999) 33–51
8. Zhou, Y., Toga, A.W.: Efficient skeletonization of volumetric objects. IEEE Transactions on Visualization and Computer Graphics **5** (1999) 196–208
9. Sonka, M., Hlavac, V., Boyle, R.: Image Processing, Analysis and Machine Vision. 2 edn. International Thomson Publishing (1998)

10. Michie, D., Spiegelhalter, D.J., Taylor, C.C.: Machine Learning, Neural and Statistical Classification. Ellis Horwood Series in Artificial Intelligence, NJ, USA (1994)
11. Lachenbruch, P.A.: Discriminant Analysis. Hafner Press, New York (1975)
12. Duda, R.O., Hart, P.E.: Pattern Classification and Scene Analysis. Wiley-Interscience Publication, New York (1973)
13. Bishop, C.M.: Neural Networks for Pattern Recognition. Oxford University Press, New York (1999)
14. Haykin, S.: Redes Neurais: Princípios e Prática. 2 edn. Bookman, Porto Alegre (2001)
15. Burges, C.J.C.: A Tutorial on Support Vector Machines for Pattern Recognition. Kluwer Academic Publishers. (1998)
16. Fukunaga, K.: Introduction to Statistical Pattern Recognition. 2 edn. Academic Press, London (1990)
17. Technologies, L.: SPSS 11.0 for windows. Available at http://www.spss.com (2003)
18. Chang, C.C., Lin, C.J.: LIBSVM – a library for support vector machines (2003) Available at http://www.csie.ntu.edu.tw/ cjlin/libsvm/.
19. Software, C.X.: Neuralpower professional v. 1.0. Available at http://www.geocities.com/neuralpower/ (2003)
20. Metz, C.E.: ROCKIT software. Available at http://www-radiology.uchicago.edu/krl/toppage11.htm (2003)
21. Greinera, M., Pfeifferb, D., Smithc, R.: Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests. Preventive Veterinary Medicine **45** (2000) 23–41