4

Prediction Models for Malignant Pulmonary Nodules Based-on Texture Features of CT Image[1]

Guo Xiuhua, Sun Tao[1], Wang huan and Liang Zhigang[2]
[1]*School of Public Health and Family Medicine, Capital Medical University, Beijing, 100069*
[2]*Department of Radiology, Xuan Wu Hospital, Capital Medical University, Beijing 100050*
China

1. Introduction

Lung cancer is one of the most harmful forms of cancer, which is the leading cause of cancer death in many regions of the world (Ahmedin Jl et al.,2005). The overall 5-year survival rate of lung cancer patients is only 14%, and remained at this level for the past two decades. However, when lung cancer is found at the early stage I or II, 5-year survival rates can be as high as 60-70% (Beadsmoore CJ et al.,2003). Early diagnosis of lung cancer was only 15%(Li YR et al., 2007). Although histology diagnosis is the most accurate detection method in the medical environment, it is an aggressive invasive procedure that involves risks, discomfort and trauma, which restrict it to be used in the clinical practice. Digital CT (Computed Tomography), overcoming the shortages of histology diagnosis, has gradually become the best imaging diagnosis method of lung cancer. CT enables us to visualize lung anatomy in great detail and has been used to accurately diagnose lung diseases since the 1980s (Ye X et al.,2006).Detecting and diagnosing solitary pulmonary nodules (SPNs, referring to the lesion of lung field ≤ 3 cm in diameter), the most common manifestation of lung cancer, are critical since early identification of malignant nodules is crucial to the chance for successful treatment. But pulmonary nodules of lung cancer in CT images share similarity with benign cases to some extent, such as tuberculosis, inflammatory pseudotumor, hamartoma, and aspergillosis(Jee WC et al.,2008), which makes it difficult to distinguish, especially for the doctors who are not rich in clinical experience. With technique of computer rising, the computer-aided diagnosis (CAD) has become an auxiliary diagnosis tool (Jiang J et al.,2007), especially in diseases that can not be diagnosed efficiently. To improve the accuracy and efficiency of CT screening programs for the detection of early-stage lung cancer, a number of

[1] Program of Funds: The program of Natural Science Fund of China (Serial Number: 30972550); the program of Natural Science Fund of Beijing (Serial Number: 7092010); the program of Academic Human Resources Development in Institutions of Higher Learning Under the Jurisdiction of Beijing Municipality（Serial Number: PHR201007112）

research projects, such as texture analysis(Liu YN et al.,2008) and image segmentation(Sun XJ et al.,2006), have been done to assist radiologists in diagnosing lung cancer.

2. Protocols of CT scan

In this study, the chest CT examinations were performed by using 64 detector-row helical CT (Cardiac-64, Siemens Medical systems, Germany) with the following parameters: 0.5s tube rotation and 1.5 pitch. A caudal-cranial direction scan was performed during an aspiratory breathhold and no contrast was used. Images were obtained from the level of the lung bases (posterior recesses) to the lung apex with the help of a scout view. Exposure settings were 150 mAs and 120 kVp. The fields-of-view were large enough to cover the complete lung cross-section. Each chest CT examination was reconstructed using two different settings immediately after imaging with the following three combinations of section thickness/increment and kernel: (A) 1.0mm/1.0 mm and a soft kernel (Siemens B30 filter), (B) 1.0 mm/1.0 mm and a sharp kernel (Siemens B60). The Siemens B30 kernel is the standard soft-tissue reconstruction kernel, and B60 is the bone reconstruction kernel, widely used in high resolution chest CT at normal. Images were displayed with a lung (level, −600 HU and width, 1500 HU) and mediastinal (level, 30 HU and width, 400 HU) window settings. Slice thickness and reconstruction intervals for routine scanning were 1-5mm.Data were reconstructed with a matrix of 512×512. Diameter range is 1.0-3.0 cm.

3. Methods of texture extraction

Nowadays, the methods of texture extraction can be classified into four parts: statistical method, model method, spectrum method and structural method. The basic procedure of texture analysis is to extract texture of images using different methods and then run a set of mathematical texture operators to produce a corresponding set of texture feature values in order to describe character of images.

Co-occurrence is one category of Statistical methods, which is a measure of the relative frequency or joint probability of two image properties occurring under predefined constraints, across the domain of an image. Gray level co-occurrence matrix (GLCM) is the most widely used texture analysis method in biological imaging (Ondimu SN et al.,2008). GLCM holds potential for analyzing segmented images of biogenic sedimentary structures because it can be used to analyze multi-scale differences in image texture (Honeycutt CE, et al.,2008). ROIs (small pulmonary nodules) were segmented using gray level threshold algorithm(Chou YC et al.,2007). Fig. 1 shows an example CT scan and a segmented slice of small pulmonary nodule. Using this segmentation algorithm, the small pulmonary nodules images were generated.

Curvelet transform, a kind of spectrum method, stems from Wavelets theory, but it overcomes the weakness of traditional multiscale representations using wavelets, and is suitable to capture more directional features in an image.

The main formulas offering to Curvelet transform are as followed:

$$\phi_{j,l,k}(X) = \phi_j(R_{\theta_1}(X - X_k^{(j,l)})) \tag{1}$$

where R_θ is the rotation by θ radians and R_θ^{-1} its inverse

Variable	Description	Formula
Energy	1 if Energy>0.20 0 otherwise	$f_1 = \sum_{i}^{M}\sum_{j}^{N} P^2(i,j)$
Inertia	1 if Inertia≤0.45 0 otherwise	$f_2 = \sum_{i=0}^{K-1}\sum_{j=0}^{K-1}(i-j)^2 c_{ij}$
Inverse Difference Moment	1 if Inverse Difference Moment >0.87 0 otherwise	$f_3 = \sum_{i}^{M}\sum_{j}^{N}\frac{P(i,j)}{\lvert i-j\rvert^k}, i \neq j$
Entropy	1 if Entropy≤0.89 0 otherwise	$f_4 = -\sum_{i=0}^{l-1}\sum_{j=0}^{l-1} p(i,j)\log_2 p(i,j)$
Correlation	1 if Correlation>0.98 0 otherwise	$f_5 = \frac{\sum_{i=0}^{l-1}\sum_{j=0}^{l-1} ijp(i-j) - \mu_1\mu_2}{{\sigma_1}^2{\sigma_2}^3}$
Cluster Tendency	1 if Cluster Tendency≤11.65 0 otherwise	$f_6 = \sum_{i}^{M}\sum_{j}^{N}(i+j-2\mu)^k P(i,j)$
Contrast	1 if Contrast≤0.45 0 otherwise	$f_7 = \sum_{i}^{M}\sum_{j}^{N}(i-j)^2 P(i,j)$
Homogeneity	1 if Homogeneity>0.88 0 otherwise	$f_8 = \sum_{i}^{M}\sum_{j}^{N}\left(\frac{P(i,j)}{1+\lvert i-j\rvert}\right)$
Variance	1 if Variance≤42.36 0 otherwise	$f_9 = \frac{1}{2}\sum_{i}^{M}\sum_{j}^{N}((i-\mu)^2 P(i,j) + (j-\mu)^2 P(i,j))$
Maximum probability	1 if Maximum probability≤0.36 0 otherwise	$f_{10} = \max_{i,j}^{M,N} P(i,j)$
Sun-mean	1 if Sun-mean≤11.66 0 otherwise	$f_{11} = \frac{1}{2}\sum_{i}^{M}\sum_{j}^{N}(iP(i,j) + jP(i,j))$
Difference-mean	1 if Difference-mean≤0.33 0 otherwise	$f_{12} = \frac{1}{2}\sum_{i}^{M}\sum_{j}^{N}(iP(i,j) - jP(i,j))$
Sum-Entropy	1 if Sum Entropy≤2.43 0 otherwise	$f_{13} = -\sum_{k=0}^{2K-2} c_{x+y}(k)\log\{c_{x+y}(k)\}$
Difference-Entropy	1 if Difference Entropy≤0.45 0 otherwise	$f_{14} = -\sum_{k=0}^{K-1} c_{x-y}(k)\log\{c_{x-y}(k)\}$

Table 1. Descriptions and formulas of fourteen Image-level texture features as variables used in the analysis

$$R_\theta = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix},\ R_\theta^{-1} = R_\theta^T = R_{-\theta}$$

A curvelet coefficient is the inner product between an element $f \in L^2(R^2)$ and a Curvelet $\phi_{j,l,k}$

$$c(j,l,k)'' := \int_{R^2} f(X)\overline{\phi_{j,l,k}(X)}dx \tag{2}$$

where R denotes the real line. Curvelet transform obeys an anisotropy scaling relation, length ≈ $2^{-j/2}$, width = 2^{-j}, so, width≈length^2. This equation called a curve scaling law.

Based on Curvelet transform, we extracted fourteen texture features of pulmonary nodules of CT images, including Entropy, Mean ,Correlation, Energy, Homogeneity, StdDev, MP, IDM, ClustTend, Inertia, SumMean, DiffMean, SumEntr, and DiffEntr. The meanings of some texture features are as follows.

Energy is defined to measure the number of repeated pairs, which is expected to be high if the occurrence of repeated pixel pairs is high. In statistical mechanics, entropy is defined as a factor or quantity that is a function of the physical state of a mechanical system and is equal to the logarithm of the probability of the occurrence of the particular molecular arrangement in that state. Inverse Difference Moment tells us about the smoothness of the image, like homogeneity. The IDM is expected to be high if the gray levels of the pixel pairs are similar. Inertia reflects the roughness of texture, which is expected to be low if the more elements are near to diagonal line of matrix when texture is rougher. Correlation is expected to measure the relevance of the gray of pixel. Sun–mean (mean) and Difference-mean provide the mean of the gray levels of the image. The sum–mean is expected to be large if the sum of the gray levels of the image is high. Standard deviation tells us how to spread out the distribution of gray levels . The variance is expected to be large if the gray levels of the image are spread out greatly. Results in the pixel pair is most predominant in the image. The Maximum probability (MP) is expected to be high if the occurrence of the most predominant pixel pair is high. The mean of the gray reflects the central tendency of the gray. Cluster tendency measures the grouping of pixels that have similar gray level values. Homogeneity measures the local homogeneity of a pixel pair. The homogeneity is expected to be large if the gray levels of each pixel are similar.

Curvelet transform is a new image representation approach that codes image edges more efficiently than wavelet transform. Curvelet will be better than wavelet in following cases (Candes EJ et al.,2006) :

1. Optimally sparse representation of objects with edges.
2. Optimal image reconstruction in severely ill-posed problems.
3. Optimal sparse representation of wave propagators.

Some studies have been done using Curvelet transform in image processing. Dettori and Semler(Lucia D et al.,2007)presented a comparative study between Wavelet, Ridgelet and Curvelet transform on some computed tomography (CT) scans. The comparative study indicated that Curvelet yields better results than Wavelet or Ridgelet.

4. Prediction models

Using texture feature values, we can establish model to predict the characteristics of pulmonary nodules. The methods of establishing prediction model are variable, such as

logistic regression, discriminant analysis, artificial neural networks, and support machine vector. Because the same patient has many CT images, that is, there is correlation among CT images of one patient. Common mathematical methods, such as logistic regression, discriminant analysis, are not appropriate to predict the characteristics of pulmonary nodules.

Multilevel modeling techniques are appropriate when there is correlation among clusters of subjects. It is the presence of within-cluster correlation that justifies the use of a multilevel (hierarchical) model, and correlation multilevel modeling without within-cluster does not provide benefit (Kim DG et al.,2007). Now we take establishing prediction model of CT images for example. The authors identified there is correlation among CT images of one patient, so multilevel models were fitted to a two-level hierarchy and used to identify factors affecting texture features of benign and malignant CT images for individual casualties. By establishing a multi-level model of texture features of pulmonary nodules, the characteristics of pulmonary nodules in the CT images could be better described , which profit early identification of small pulmonary nodules.

We make small pulmonary nodules CT images as level 1 and SPN patients as level 2. With two-level structure data, three different equations can be formulated: individual-level model(image－level model, level 1 model), organization-level model(patient-level model, level 2 model), and combined model. Assuming normally distributed errors, for subject ij we have level 1 model, level 2 model and combined model (Wolfinger R et al.,1993), as

$$Y_{ij} \sim N(\hat{Y}_{ij}, \sigma^2{}_{ij});\ r_{ij} \sim N(0, \sigma^2);\ \hat{Y}_{ij} = \hat{\beta}_{0j} + \hat{\beta}_{1j} X_{ij};$$

$$Y_{ij} = \beta_{0j} + \beta_{1j} X_{ij} + r_{ij} \quad \text{(level 1 model)} \tag{3}$$

$$\beta_{oj} = \gamma_{00} + \gamma_{01} W_j + \mu_{0j} \quad \text{(level 2 model)} \tag{4}$$

$$\text{and } \beta_{1j} = \gamma_{10} \quad \text{(level 2 model)} \tag{5}$$

Substituting Eqs.(4)and(5) into Eq.(3) yields the combined model:

$$Y_{ij} = \gamma_{00} + \gamma_{01} W_j + \gamma_{10} X_{ij} + \mu_{oj} + r_{ij} \quad \text{(combined model)} \tag{6}$$

If the observed outcomes Y_{ij} are binary, a binomial logistic model is appropriate. A multilevel binomial logistic model for outcome probabilities of benign and malignant pulmonary nodules on CT image data used in this study is formulated as follows:

$$\log(\frac{p_{ij}}{1-p_{ij}}) = \gamma_{00} + \sum_{q=1}^{Q} \gamma_{0q} W_{qj} + \sum_{p=1}^{P} \gamma_{p0} X_{pij} + u_{0j} \tag{7}$$

Where P is the probability that malignant pulmonary nodules on CT image will occur (Y_{ij} =1), γ_{00} the intercept, W_{qj} a vector of patient-level characteristics, X_{pij} a vector of image-level characteristics, and the regression coefficients associated with the patient-level characteristics and the image-level characteristics, respectively, and u_{0j} is the random effect at level 2,where $u_{0j} \sim N(0, \sigma_u{}^2)$.

SVM is a popular classifier based on structural risk minimization principle (Vapnik VN.,1998), which could minimize the generalization error of the classifier. Recently, SVM has gained much attention as a useful tool for image recognition. Youngjoo Lee(Youngjoo L et al.,2009) investigated the performance of Bayesian classifier, ANN (artificial neural net) and SVM (support vector machine) for differentiating obstructive lung diseases using texture analysis. Results showed that SVM showed the best performance for classification. The same result had been got by Michael E. Mavroforakis(Michael EM et al.,2006) .

Compared with other classifiers, such as Artificial Neural Networks, SVM aims to find the hyperplane that maximizes the distance from the hyperplane to the nearest examples in each class. An attractive feature of SVM is that it can map linearly inseparable data into higher dimensional space so that SVM can make them to be linearly separable. There are two types of SVM, linear and non-linear. The training data of linear SVM may be analyzed as either linearly separable or linearly non-separable. Given a set of training vectors (l in total) belonging to separate classes (x_1,y_1), (x_2,y_2), (x_3,y_3), … ,(x_1,y_1) , where $x_j \in R^n$ denotes the ith input vector and $y_j \in \{+1,-1\}$ is the corresponding desired output. The maximal margin classifier aims to find a hyperplane w: wx+b=0 to separate the training data. In the possible hyperplanes, only one maximizes the margin (distance between the hyperplane) and the nearest data point of each class. The support vectors denote the points lying on the margin border (Huang YL,2005). The solution to the classification is given by the decision function

$$f(x) = sign\left(\sum_{j=1}^{N_{sv}} \alpha_j y_j (s_j, x) + b\right) \tag{8}$$

where α_j is the positive Lagrange multiplier, s_j is the support vectors (N_{sv} in total), and $k(s_j,x)$is the function for convolution of the kernel of the decision function. Such kernels must hold Mercer´s condition(V V,1982) which tells us whether or not a perspective kernel is a dot product in some space. The polynomial, radial, anova kernels are now often seen choices in SVM-based CAD applications.

5. Examples

In the rest of the paper, we will provide two practical examples to explain the use of prediction model for small pulmonary nodules, which based on texture extraction to predict the characteristics of pulmonary nodules.

5.1.Example1: Multilevel binomial logistic prediction model for malignant pulmonary nodules based on texture features of CT image.

The digitized CT image set used in this study contains 2171 ROIs (Region of Interests) extracted from 185 patients with small solitary pulmonary nodules, with 61 benign nodules and 124 malignant tumors. There were 107 men and 78 women (range of age, 19-80 years; mean ages, 58 years). The final diagnosis of 124 small peripheral lung cancers (diameter range, 1.0-3.0 cm; mean diameter, 2.0cm) was determined by either operation or biopsy. All the images were provided by the radiology department of Beijing Friendship Hospital affiliated to Capital University of Medical Science.

The structure of data from 185 patients is postulated as hierarchical data, which consists of two different levels: level-1 consisting of image-level characteristics and level-2 consisting of

patient-level characteristics. Image-level characteristics contain detailed information associated with individual images such as Energy, Contrast, and Inverse Difference Moment and the patient-level characteristics include sex and age. Fourteen image-level and two patient-level variables are used as independent variables in the analysis, and the benign and malignant pulmonary nodules as the dependent variables, 1 malignant, and 0 benign. Sex and age are patient-level variables (1 man, 0 woman ; 1 age >50.00 0 others). The descriptions of the fourteen image-level variables used in the study are provided in Table 1. Besides, Table 1 gives formulas of fourteen GLCM textural features in the study (Dettori L et al.,2007; Yogesan1 K et al.,1996; Guo XH et al.,2008).

In this example, we used gray level co-occurrence matrix to get fourteen textural features and establishd multilevel binomial logistic prediction model (Wang H et al.,2010). combining patient and image characteristics of textural features Results showd that Five texture features, including Inertia, Entropy, Correlation, Difference-mean, Sum-Entropy, and age of patients own aggregating character on patient-level, were statistically different (P<0.05) between benign and malignant small solitary pulmonary nodules.

For multilevel binomial logistic models, the variance at the lowest level is completely determined by the population proportion (Kim DG et al.,2007). SAS software (version 9.1, SAS Institute (Shanghai) Co., Ltd.) was used to perform the estimation of multilevel binomial logistic models.

For obtaining estimates of between- and within-organization (or cluster) variance, null models were estimated (Table 2). The intra-class correlation coefficient (ICC) is 0.1795 for CT images, indicating that 17.95% of the total variation in images exists between patients, and therefore may be explained using patients-level predictors. As a result, the patients-level predictors are useful for estimating statistical models for texture features of CT images. In other words, multilevel models for texture features of CT images are necessary. It should be noted that roughly 18% of the total variation in texture features of CT images is attributable to the variability between patients, which suggesting that texture features of CT images are significantly influenced by patient's characteristics.

	external segmentation
Fixed effect	
Intercept	0.6766 (0.0100)
Random effect	
Images-level	0.03928(0.0066)
Patients-level	0.1795 (0.0000)
ICC	0.1795
-2 Log Likelihood	2861.9

Note. For parameter estimates, standard errors appear in parentheses.

Table 2. The estimation results of null models

Based on the results of null model estimation, one binomial logistic regression model and multilevel binomial logistic regression model can be used to estimate the texture features of CT images.

Table 3 presents the estimation results of CT images model, in which image and patient features are included as predictors. For logistic regression models, the odds ratio is used to interpret the actual effects of estimated coefficients. Odds ratios are also provided in Table 3. The results show that malignant pulmonary nodules in CT image are more likely to occur while Inertia is lower than 0.4435 (odds=1.494–1), Difference-mean is lower than 0.3315 (odds=1.332–1) or Inverse Difference Moment is higher than 0.8662 (odds=1.156–1) compared to benign pulmonary nodules. The results also show that malignant pulmonary nodules in CT image are less likely to occur while Entropy is lower than 0.8939 (odds =0.757–1), Sum- Entropy is lower than 2.4314 (odds =0.877–1) or Correlation is higher than 0.9754 (odds = 0.779–1) compared to benign pulmonary nodules. Malignant pulmonary nodules in CT image belongs to young patients (≤50) are less likely (odds= 0.503–1) than old patients (>50). These findings are consistent with warrants for old patients of the effects of small solitary pulmonary nodules. That means old patients are 49.7% ((1-0.503) ×100) more likely to get earlier period lung cancer than young patients. The sensitivity of multilevel binomial logistic prediction model was 90.6% for another 50 patients with small solitary pulmonary nodules, which had a good effect on prediction of small pulmonary nodules. The result of prediction would be improved with the enhancement of doctors' clinical experience.

5.2 Example 2: Support vector machine prediction model for small pulmonary nodules based on Curvelet transform to extract texture features of CT image

In this example, we explore the use of Curvelet transform to extract texture features of pulmonary nodules in CT image and support vector machine to establish prediction model of small solitary pulmonary nodules in order to promote the ratio of detection and diagnosis of early-stage lung cancer. Results show that the classification consistency, sensitivity and specificity for the model are 81.5%, 93.8% and 38.0% respectively.

2461 CT images used in this study are extracted from 129 patients with small solitary pulmonary nodules, including 537 CT images (25 benign cases) related to benign nodule and 1924 CT images (104 malignant cases) to malignant tumors. The final diagnosis of malignant cases was determined by either operation or biopsy. The diagnosis of benign cases was confirmed by operation, CT diagnosis or follow-up. The original format is DICOM, and diameters of the chest nodules were from 0.3 cm to 3 cm. 129 cases were provided by four hospitals, and details are as follows: Beijing Xuanwu Hospital of Capital Medical University (26 malignant cases, 11 benign cases), Beijing Friendship Hospital affiliated to Capital Medical University (35 malignant cases,6 benign cases), Chaoyang Hospital affiliated to Capital Medical University (20 malignant cases,7 benign cases) and Fuxing Hospital affiliated to Capital Medical University (23 malignant cases,1 benign cases).

Based on Curvelet transform, we extracted fourteen texture features of pulmonary nodules of CT images. Every image could be decomposed into 18 sub-images. The 18 sub-images could be classified into three parts: inner layer, middle layer and outer layer. So 252 texture features were extracted from every image. Among those texture features, 158 texture features showed statistically significant differences between benign and malignant cases

	Estimate	Odds ratio	95% Confidence Limits
Fixed effects			
Intercept (γ_{00})	-0.0638 (0.1204)	0.9382	(-0.2997 0.1721)
Image-level			
Energy (γ_{10})	0.0776 (0.0683)	1.0807	(-0.0562 0.2114)
Inertia (γ_{20})	0.4014*** (0.1316)	1.4940	(0.1434 0.6594)
Inverse Difference Moment (γ_{30})	0.1450* (0.0813)	1.1560	(-0.0143 0.3043)
Entropy (γ_{40})	-0.2779*** (0.0603)	0.7574	(-0.3960 -0.1597)
Correlation (γ_{50})	-0.2493 *** (0.0956)	0.7793	(-0.4366 -0.0620)
Cluster Tendency (γ_{60})	0.0174 (0.0631)	1.0176	(-0.1062 0.1410)
Contrast (γ_{70})	-0.0461 (0.0743)	0.9549	(-0.1919 0.0996)
Homogeneity (γ_{80})	0.0904 (0.1425)	1.0946	(-0.1889 0.3696)
Variance (γ_{90})	0.0971 (0.0676)	1.1020	(-0.0353 0.2296)
Maximum probability (γ_{100})	0.1098 (0.0686)	1.1161	(-0.0247 0.2443)
Sun-mean (γ_{110})	0.0174 (0.0631)	1.0176	(-0.1062 0.1410)
Difference-mean (γ_{120})	0.2863** (0.1386)	1.3315	(0.0146 0.5580)
Sum-Entropy (γ_{130})	-0.1311** (0.0648)	0.8771	(-0.2581 -0.0041)
Difference-Entropy (γ_{140})	-0.1755 (0.1595)	0.8390	(-0.4881 0.1370)
Patient-level			
sex(γ_{01})	0.0781 (0.0581)	1.0812	(-0.0359 0.1920)
age(γ_{02})	-0.6871*** (0.0611)	0.5030	(-0.8069 -0.5674)
Random effects			
τ_{00} (μ_{0j})	0.4583*** (0.0280)		

Note. For parameter estimates, standard errors are within parentheses. *P< 0.10; **P< 0.05; ***P< 0.01

Table 3. Estimation results for CT images

through two independent samples tests of nonparametric test or two independent samples t-test

The 2461 images were divided into two parts: one part was as a training sample (80%) and the other part was as a test sample (20%). The training sample was used to establish the database and the test sample was used to evaluate the validity of prediction model of SVM (Table 4).

Samples	Benign	Malignant	Total
Training sample	429	1539	1968
Test sample	108	385	493
Total	537	1924	2461

Table 4. Benign and Malignant Cases Distribution

Based on Curvelet transform, 252 texture features we extracted were as parameters to establish prediction model for small pulmonary nodules (Table 5).

SVM	Pathological Diagnosis		Total
	Benign	Malignant	
Benign	41	24	65
Malignant	67	361	428
Total	108	385	493

Table 5. Prediction Results of Pulmonary Nodules Based On SVM

The validity of prediction model of SVM is evaluated by the following three indexes: sensitivity (93.8%), specificity (38.0%) and consistency (81.5%). The high sensitivity (93.8%) can reduce the false negative rate of early-stage lung cancer effectively.

There are other methods used in published papers to select texture features. Wavelet transform was used to extract the texture features of chest radiography, and the Energy was as the only parameter to establish the prediction model (Huang PW. et al.,2004). Lucia Dettori(Lucia D et al.,2007) selected Mean, StaDev, Energy and Entropy to establish the prediction model. Principal component analysis，a very useful tool to deal with colinearity,

has various applications in texture extraction and tumor recognition(Zhang J et al.,2008). Mohamed Meselhy Eltoukhy(Mohamed ME et al.,2010) used Curvelet transform to decompose mammogram images into 4 levels, then selected the largest 100 texture features as parameters.

In order to select texture features which are more accurate to reflect characteristics of pulmonary nodules, we have made many attempts. Results were showed in table 6.

In order to promote sensitivity and specificity, we had made some attempts to select proper texture features. Compared with other methods, 252 texture features were used as parameters to establish prediction model is more satisfying.

Based on published reports, characteristics of pulmonary nodules can been detected by texture features. However, 2D images are irregular when decomposed, and the Curvelet transform is more suitable than the wavelet transform to extract texture features. The methods to establish prediction model are variable, such as multiple linear regression, logistic regression, discriminant analysis, artificial neural networks, but the result of support vector machine is better (Zheng Z et al.,2007).In this research, we establish support vector machine prediction model for small pulmonary nodules using Curvelet transform to extract texture features of CT image, which has not been reported to our knowledge.

	Sensitivity(%)	Specificity(%)	Consistency(%)
Using Energy As The Only Parameter	93.2	29.6	79.3
Using Texture Features of Inner Layer As Parameters	96.4	31.5	82.2
Using Texture Features of Middle Layer As Parameters	94.8	25.0	79.5
Using Texture Features of Outer Layer As Parameters	100.0	0.0	78.1
Using Mean, StaDev, Energy and Entropy As Parameters	94.8	29.6	80.5
Using Principal Component Analysis	100.0	0.0	78.1
Using 158 Texture Features As Parameters	94.5	34.3	81.3
The Largest 100 Texture Features As Parameters	93.8	28.7	79.5

Table 6. Prediction Results of Pulmonary Nodules Using Other Methods

6. Summary

In recent years, the incidence of lung cancer has been the top of cancers in the most countries. Because of the difficulty to diagnosis, more attention has been paid to lung cancer.

Now the most accurate diagnosis method of lung cancer is histology diagnosis, but this method is traumatic, which restricts it to be used in clinical practice. In the decades, digital CT has been the main diagnosis tool of lung cancer for its convenience and safety, and widely used in clinical practice. However, it is difficult to distinguish between benign and malignant cases in the CT images of pulmonary nodules, especially for the doctors who were lack of experience. From two examples, we can make the conclusion that the prediction model is so sensitive that it can diagnose early-stage lung cancer effectively, reduces the difficulty of distinguishing characteristics of pulmonary nodules and improves accuracy rate of diagnosing early-stage lung cancer.

7. References

Ahmedin Jl, DVM, PhD, Taylor M, Elizabeth W, PhD, Alicia S, MPH, Ram CT, PhD, Asma G, MPH, Eric JF, PhD, MJT, MD, MS. (2005). Cancer statistics 2005, A Cancer Journal for Clinicians Vol.55(No.1):10-30.

Beadsmoore CJ, Screaton NJ. (2003). Classification, staging and prognosis of lung cancer, European Journal of Radiology Vol.45(No.1):8-17.

Chou YC, Teng MM, Guo WY, Hsieh JC, Wu YT. (2007). Classification of hemodynamics from dynamic susceptibility contrast magnetic resonance (DSC-MR) brain images using noiseless independent factor analysis, Medical Image Analysis Vol.11(No.3): 242–253.

Dettori L, Semler L. (2007). A comparison of wavelet, ridgelet, and curvelet-based texture classification algorithms in computed tomography, Computers in Biology and Medicine Vol.37(No.4): 486－498.

Guo XH, Zhang Y, Wang H, Li KC, Yao XY, Liang ZG.(2008). Exploring the Risk Factors of Lung Pulmonary Nodules with Cancer Using Sandwich Logistic Regression Analysis, Journal of Mathematical Medicine Vol.22(No.1): 66-68.

Honeycutt CE, Plotnick R. (2008) Image analysis techniques and gray level co-occurrence matrices (GLCM) for calculating bioturbation indices and characterizing biogenic sedimentary structures, Computers & Geosciences.

Huang PW., Dai SK. (2004). Design of a two-stage content-based image retrieval system using texture similarity, Information Processing & Management Vol.40 (No.1): 81–96.

Jee WC, Chin AY, Dae-Soon S, Naeyun C, Jinseon L, Hong KK, Yong SC, Kyung SL, Jhingook K.(2008). Prediction of lymph node metastasis using the combined criteria of helical CT and mRNA expression profiling for non-small cell lung cancer, Lung Cancer.

Jiang J, Yao B, Wason AM. (2007). A genetic algorithm design for micro calcification detection and classification in digital mammograms, Computerized Medical Imaging and Graphics Vol.31(No.1):49–61.

Kim DG, Lee Y, Simon W, Keechoo C. (2007). Modeling crash outcome probabilities at rural intersections: Application of hierarchical binomial logistic models. Accident Analysis and Prevention Vol.39(No.1):125-134

Li YR, Yang XF. (2007) The Progress on Clinical Early Diagnostic Methods of Lung Cancer, Journal of Clinical Pulmonary Medicine Vol.12(No.2):130-132.

Liu YN, Wang H, Guo XH, Liang ZG, He Q. (2008). Application of artificial neural networks in prediction model of early-stage lung cancer, Chinese journal of Medical Statistics Vol.1(No.1): 30–33.

Lucia D, Lindsay S. (2007). A comparison of wavelet, ridgelet, and curvelet-based texture classification algorithms in computed tomography, Computers in Biology and Medicine. 486-498.

Michael EM, Harris VG, Nikos D, Dionisis C, Sergios T. (2006). Mammographic masses characterization based on localized texture and dataset fractal analysis using linear, neural and support vector machine classifiers, Artificial Intelligence in Medicine. 145-162.

Mohamed ME, Ibrahima F, Brahim BS. (2010). A comparison of wavelet and curvelet for breast cancer diagnosis in digital mammogram, Computers in Biology and Medicine. 384–391.

Ondimu SN, Murase H. (2008). Effect of probability-distance based Markovian texture extraction on discrimination in biological imaging, Computers and Electronics in Agriculture Vol.63(No.1): 2-12.

Sun XJ, Zhang HB, Duan HC. (2006). 3D computerized segmentation of lung volume with computed tomography, Academic Radiology Vol.13(No.6):670–677.

Vapnik VN. (1998). Statistical learning theory. New York: Wiley.

V V. (1982). Estimation of Dependencies based on Empirical Data, Springer Verlag, New York.

Wang H, Guo XH, Jia ZW, Li HK, Liang ZG, Li KC, He Q. (2010). Multilevel binomial logistic prediction model for malignant pulmonary nodules based on texture features of CT image. European Journal of Radiology Vol.74(No.1): 124-129.

Wolfinger R, O'connell M. (1993). Generalized linear mixed models: a pseudolikelihood approach, Journal of Statistical Computation and Simulation Vol.488(No.3&4):233–243.

Ye X, Edwin J.R, Yu H, Guo JF, Geoffrey M, Eric AH. (2006). Computer-aided Classification of Interstitial Lung Diseases Via MDCT: 3D Adaptive Multiple Feature Method (3D AMFM), Academic Radiology Vol.13(No.8):969-978.

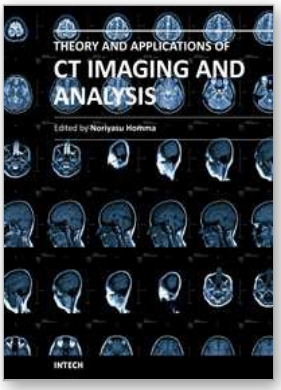
Yogesan1 K, Jørgensen T, Albregtsen F, Tveter KJ, Danielsen HE. (1996). Entropy-Based Texture Analysis of Chromatin Structure in Advanced Prostate Cancer, Cytometry Vol.24(No.3):268-276.

Youngjoo L, Joon BS, June GL, Song SK, Namkug K, Suk HK. (2009). Performance testing of several classifiers for differentiating obstructive lung diseases based on texture analysis at high-resolution computerized tomography (HRCT), Computer Methods And Programs In Biomedicine. 206-215.

Huang YL, Chen DR. (2005). Support vector machines in sonography Application to decision making in the diagnosis of breast cancer, Journal of Clinical Imaging.179–184

Zhang J, Tong LZ, Lei W, Ning L. (2008). Texture analysis of multiple sclerosis: a comparative study, Magnetic Resonance Imaging. 1160–1166.

Zheng Z, Zhang YX, HU YX. (2007). Investigation of eye gaze based on independent component analysis and support vector machine[J], Journal of Optoelectronics.Laser Vol.18(No.7): 491-494.



Theory and Applications of CT Imaging and Analysis
Edited by Prof. Noriyasu Homma

ISBN 978-953-307-234-0
Hard cover, 290 pages
Publisher InTech
Published online 04, April, 2011
Published in print edition April, 2011

The x-ray computed tomography (CT) is well known as a useful imaging method and thus CT images have continuingly been used for many applications, especially in medical fields. This book discloses recent advances and new ideas in theories and applications for CT imaging and its analysis. The 16 chapters selected in this book cover not only the major topics of CT imaging and analysis in medical fields, but also some advanced applications for forensic and industrial purposes. These chapters propose state-of-the-art approaches and cutting-edge research results.

How to reference
In order to correctly reference this scholarly work, feel free to copy and paste the following:

Guo Xiuhua, Sun Tao, Wang huan and Liang Zhigang (2011). Prediction Models for Malignant Pulmonary Nodules Based-on Texture Features of CT Image, Theory and Applications of CT Imaging and Analysis, Prof. Noriyasu Homma (Ed.), ISBN: 978-953-307-234-0, InTech, Available from: http://www.intechopen.com/books/theory-and-applications-of-ct-imaging-and-analysis/prediction-models-for-malignant-pulmonary-nodules-based-on-texture-features-of-ct-image

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821