

A Comparative Study of HRCT Image Metrics and PFT Values for Characterization of ILD and COPD

Gang Song¹, Nicholas Tustison¹, Eduardo Barbosa Jr², James C. Gee¹, Warren Gefter², Kreider Mary³, Drew Torigian²

¹ Penn Image Computing and Science Laboratory (PICSL), Dept. of of Radiology, University of Pennsylvania, Philadelphia, PA, USA,

² Dept. of Radiology, University of Pennsylvania, Philadelphia, PA, USA

³ Dept. of Medicine, University of Pennsylvania, Philadelphia, PA, USA

Abstract. Several image metrics have been proposed for pulmonary assessment via thoracic high-resolution computed tomography (HRCT) for various pathologies. This paper describes a systematic analysis of the utility of such metrics for characterizing interstitial lung disease (ILD) and chronic obstructive pulmonary disease (COPD), in comparison to data from pulmonary function testing (PFT). HRCT inspiratory and expiratory images for 14 patients with ILD and 11 patients with COPD were acquired retrospectively. PFT values were also acquired retrospectively for each patient. Using a statistical feature selection scheme, our study demonstrates that the quantitative image features perform quite well in comparison with the clinically-used PFT values. In the first 25 selected features out of the total 114 mixed image metrics and PFT values, 21 are from the image metrics. The classification using mixed selected image features and PFT values outperforms using PFT values alone. Our study also shows that these image metrics are not redundant with respect to the PFT values for characterization of ILD and COPD.

1 Introduction

Chronic lung disease constitutes a major worldwide public health care problem and is the fourth leading cause of morbidity and mortality in the United States. Based on clinical, imaging and pathological characteristics, most can be grouped within two basic categories: interstitial lung disease (ILD) and chronic obstructive pulmonary disease (COPD). ILD is a heterogeneous group of diseases in which the hallmark is chronic, progressive, predominantly interstitial inflammation with varying degrees of fibrosis of the lung parenchyma, eventually leading to reduced lung volume, decreased lung compliance and restrictive physiology. COPD is characterized by chronic airflow limitation due to small airway disease and parenchymal destruction which is not fully reversible and is usually progressive.

The diagnosis, differentiation, and classification of the severity of ILD and COPD rely on clinical assessment, thoracic imaging (using CT and chest radiography), and pulmonary function testing (PFT). PFT is a noninvasive method of

assessing the integrated mechanical function of the lung, chest wall and respiratory muscles. It currently comprises the gold standard for pulmonary assessment. Using PFT, the heterogeneous group of ILD typically exhibits a restrictive physiology pattern whereas COPD manifests as an obstructive physiology pattern.

PFT strictly permits a global assessment of lung functionality. In contrast, HRCT image analysis is a powerful tool with the potential for regional as well as global quantitation of pulmonary diseases. Although generally effective, radiologic interpretation of HRCT is time-consuming, largely qualitative, and prone to diagnostic variability. As a result, various CT image metrics have been proposed towards a more consistent and facilitated assessment. Early investigations into CT lung analysis employed relatively simple metrics, such as the mean attenuation value and other such first-order statistical measurements obtainable from the attenuation histogram [1–3]. More sophisticated metrics relying on texture descriptions of the parenchyma have shown promise in recent studies [4].

PFT values are used clinically to diagnose ILD and COPD. With the increasing amount of proposed image metrics, research inquiry concerns the effectiveness of these metrics compared with the gold standard PFT values. Previous research ([5–9]) investigated the correlation between various quantitative image metrics with different PFT values.

Instead of a correlation study, this paper addresses the question more from the view of feature selection. We put more emphasis on what image metrics and what PFT values can characterize ILD and COPD in a quantitative framework. Image metrics and PFT values are viewed as candidates for selecting which best characterize the corresponding diagnosis (i.e. ILD or COPD in this paper). We are interested in whether and what features from image metrics have additional information for diagnosis when PFT values are provided. This also differs from the classification work of [4] that we do not tend to train any classifiers directly. The selected features can be used as inputs for any available classifier. We use Support Vector Machine (SVM, [10]) as an example to test the efficiency of the feature selection results.

A minimal-redundancy-maximal-relevance (mRMR) information framework was introduced in [11] for such a feature selection task. The ideal selected features satisfy two constraints: maximal relevance and minimal redundancy. The *relevance* of both image and PFT features concerns the ability of such features in matching an existing classification (in our case, from clinical diagnosis). It is usually computed in terms of mutual information, correlation, or statistical tests. However, in order to get a compact subset of features to classify different types of disease, it is not enough to consider only the features with highest relevance. The selected features need to be as independent to each other as possible. This is known as the criterion of minimal *redundancy*.

In this manuscript, we provide a systematic relevancy/redundancy analysis comparing 31 various statistical image metrics and 21 PFT values obtained in patients with diagnosed ILD and COPD. The analysis framework is described in Section 2. In Section 3, we provide the results of our analysis in comparing the

characterization performance of both image and PFT features. This is followed by discussion of the results.

2 Materials and Methods

14 patients with ILD and 11 patients with COPD were retrospectively identified. Every patient underwent both thoracic HRCT image acquisition and PFT within 3 days of each other. HRCT was performed for both inspiration and expiration on a 64 multidetector row CT scanner (Siemens Medical Solutions) with reconstruction of contiguous 1 mm axial images with a B41f kernel. Inspiratory and expiratory image datasets were then analyzed through a computational software developed in our laboratory which is capable of generating several hundred distinct metrics encompassing various aspects of lung physiology (e.g. pulmonary volumetric and gross tissue indices, attenuation histogram statistics, deformation indices, co-occurrence [12] and run-length [13] matrix texture indices, and attenuation mask indices), gleaned from the relevant literature. For this study, we only focused on a portion of these metrics .

Our whole pipeline (Fig. 1) is fully automatic. First, a segmentation algorithm [14] is applied to extract lung regions and segment the trachea from CT images. Next, we compute 31 different types of image metrics (Table.2, gleaned from the relevant literature) on the expiratory and inspiratory image datasets separately. We also subtract expiratory metrics from their inspiratory counterparts to obtain additional 31 metrics. This translates into a total of 93 image metrics for each patient. We then compute the relevance of each PFT value and each image metric to the corresponding disease type. The criteria of mRMR [11] is applied to select a mixed subset of PFT values and image metrics to show the degree of redundancy between them. Finally, we train SVM classifiers using the selected feature set to test the accuracy of ILD/COPD classification.

2.1 PFT Parameters and Image Metrics

Table 1 lists all the 21 PFT values used in the study. These parameters were either directly obtained or calculated for each patient, using standard protocols following guidelines of the American Thoracic Society. As mentioned previously, 93 image metrics (shown in Table 2) were generated for each patient. We use

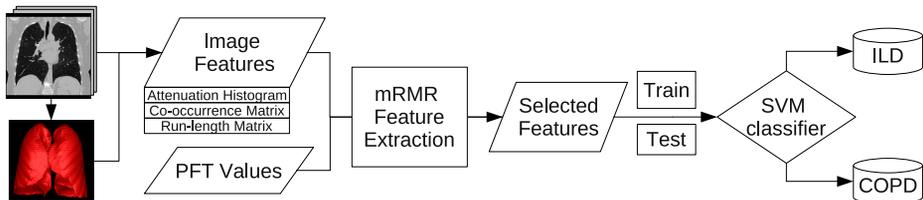


Fig. 1. Pipeline of the proposed feature selection analysis.

PFT	<i>FVC</i>	<i>FEV1</i>	<i>FEV1</i> <i>/FVC</i>	<i>FEF</i> <i>max</i>	<i>FEF 25-</i> <i>75%</i>	<i>FEF50</i>	<i>FIF50</i>	<i>MVV</i>	<i>SVC</i>	<i>IC</i>	<i>ERV</i>
<i>Rel</i>	0.088	0.026	0.47	0.357	0.218	0.307	0.021	0.251	0.197	0.083	0.146

PFT	<i>TGV</i>	<i>RV</i> <i>(Pleth)</i>	<i>TLC</i> <i>(Pleth)</i>	<i>RV</i> <i>/TLC</i>	<i>DLCO</i> <i>unc</i>	<i>DLCO</i> <i>cor</i>	<i>DL/VA</i>	<i>VA</i>	<i>Raw</i>	<i>sGaw</i>	
<i>Rel</i>	0.268	0.268	0.35	0.218	0.041	0.15	0.066	0.269	0.101	0.074	

Table 1. All 21 PFT values (1st and 3rd rows) and corresponding relevance values (2nd and 4th rows) to the disease types.

$G1$ to denote the 31 features from each of the expiratory image, $G2$ for the inspiratory images and $G3$ for the difference of the expiratory/inspiratory metric pairs. This subset of image metrics includes those first-order statistical measures generated from the attenuation histogram. It also includes more sophisticated second-order statistical quantities related to the texture of lung parenchyma, i.e. the co-occurrence [12] and run-length [13] matrix texture indices.

2.2 Feature Selection

The primary motivation for this study is to determine how image metrics perform relative to PFT values in matching clinical diagnosis of ILD and COPD. We also select an optimal subset of mixed image metrics and PFT values under the principle of minimal redundancy and maximal relevance.

In accordance with the specifications of the algorithm in [11], each of the $N = 25$ patients is given a label c based on the clinical diagnosis, either ILD or COPD. Such a disease type c is a random variable. Also each patient has a set of $J = 93$ computed image metrics $X = \{x_1, \dots, x_J\}$ and a set of $K = 21$ measured PFT values $Y = \{y_1, \dots, y_K\}$. Let the mixed feature set $Z = X \cup Y$. Each element $z_i \in Z$ is also viewed as a random variable. [11] proposed to use mutual information to measure the *relevance* between z_i and c : $I(z_i, c) = \iint p(z_i, c) \log \frac{p(z_i, c)}{p(z_i)p(c)} dz_i dc$. The mRMR framework in [11] selects a subset of S from Z such that:

$$S = \arg \max_{S \subseteq Z} \left\{ \sum_{z_i \in S} I(z_i, c) - \frac{1}{\|S\|} \sum_{z_i, z_j \in S} I(z_i, z_j) \right\} \quad (1)$$

The first term in Equ. 1 maximizes the total *relevance* of the selected features with the corresponding disease types; the second term minimizes the total redundancy of all pairs of the selected features. The framework in [11] gives a heuristic way to optimize Equ.1. We use the online toolbox [18] to compute the relevance weights and select features.

Given the diagnosis c , we analyze the relevance of each feature, which indicates its individual characterization power. Then we use mRMR to select the subset of the first 25 features. If the subset includes only PFT values, this indicates that image metrics are redundant for differentiating ILD and COPD;

	No.	Metrics Type	$G1$	$G2$	$G3$
Attenuation Histogram Indices	1	lung region volume [15][5]	0.325	0.276	0.073
	2	relative volume _{<-910} [5]	0.439	0.454	0.071
	3	attenuation mean [2, 3]	0.416	0.407	0.181
	4	attenuation variance [16]	0.171	0.197	0.181
	5	sum [2]	0.295	0.353	0.005
	6	attenuation skewness [4]	0.276	0.159	0.026
	7	attenuation kurtosis [4]	0.392	0.201	0.079
	8	attenuation grey level entropy [4]	0.021	0.121	0.005
	9	5% attenuation value [1]	0.47	0.527	0.353
	10	95% attenuation value [1]	0.463	0.22	0.006
	11	5% attenuation mean [1]	0.488	0.416	0.036
	12	95% attenuation mean [1]	0.055	0.065	0.071
Co-occurrence Matrix Indices	13	energy [4, 12]	0.019	0.05	0.05
	14	entropy [4, 12]	0.159	0.215	0.159
	15	correlation [4, 12]	0.003	0.001	0.002
	16	inverse difference moment [4, 12]	0.021	0.121	0.025
	17	inertia [12]	0.003	0.022	0.005
	18	cluster shade [12]	0.083	0.021	0.074
	19	cluster prominence [12]	0.113	0.074	0.104
20	Haralick's correlation [12]	0.034	0.034	0.001	
Run-length Matrix Indices	21	short run emphasis [4]	0.049	0.002	0.002
	22	long run emphasis [4]	0.074	0.002	0.002
	23	grey level non-uniformity [4]	0.243	0.316	0.007
	24	run-length non-uniformity [4]	0.325	0.353	0.006
	25	run percentage [4]	0.157	0.101	0.101
	26	low grey level run emphasis [17]	0.285	0.157	0.113
	27	high grey level run emphasis [17]	0.005	0.114	0.005
	28	short run low grey level emphasis [13]	0.285	0.157	0.058
	29	short run high grey level emphasis [13]	0.005	0.114	0.005
	30	long run low grey level emphasis [13]	0.285	0.157	0.113
	31	long run high grey level emphasis [13]	0.005	0.114	0.005

Table 2. Computed image metrics and their relevance to the disease types. $G1$ denotes the metrics computed from the expiratory images; $G2$ from the inspiratory images. Metrics of $G3$ are generated by subtracting $G1$ from $G2$. Within each of $G1/G2/G3$, the metrics are indexed from 1 to 31. The numbers in last three columns are the relevance values to the disease types of ILD and COPD.

otherwise, it indicates that image metrics have extra information that PFT values do not have for differentiating ILD and COPD.

2.3 Classification with SVMs

The last step in our analysis pipeline is to use the selected features to train classifiers to classify ILD and COPD. An efficient feature selection scheme should achieve low classification error rate with minimal number of features. We are going to study the relationship between classification error and the number of

the selected features. Also we are going to compare adding image metrics to PFT values as additional features to classification using PFT values only.

We use Support Vector Machines (SVM)[10] as classifiers in this paper. SVMs have been used widely for feature classification tasks. SVM classifies data by maximizing the margins between hyperplanes that separates data points of two classes. By choosing different kernels ([19]), data points are implicitly mapped to different higher dimensional spaces and thus better classification can be achieved. More discussion about SVMs can be found in [19].

The experiments here use SVM with Gaussian kernels to compare the selected features. We also test the linear kernel to understand whether the performance improvement is from the choice of nonlinear kernel or from the more efficient features.

3 Results

We first compare the relevance of each feature to the diagnosis of diseases. Fig. 2 graphically depicts the relevance weights of all the image metrics. The relevance weights of PFT values are plotted in Fig. 3 as a comparison. For example, the ratio of the forced expiratory volume in 1 second to the forced vital capacity ($FEV1/FVC$) has a high relevance value (0.47), as it is a meaningful clinical indication.

The relevance values from image metrics and PFT values are in a similar range. The top PFT parameter, $FEV1/FVC$, is 0.47; the top image metric, 5% attenuation value from the inspiratory images, is 0.52. This shows that some image metrics are as good indicators as PFT values for characterization of ILD and COPD. It also shows in Fig. 2 that the metrics computed from expiratory images ($G1$: the red bars) are similar to the metrics from inspiratory images ($G2$: the blue bars). Their relevance values have a high correlation of 0.84.

We use mRMR to select a subset of 250 features. As discussed before, a high relevance does not suffice for a feature to be selected. It also requires a selected feature to have low redundancy with respect to other features in the subset. In

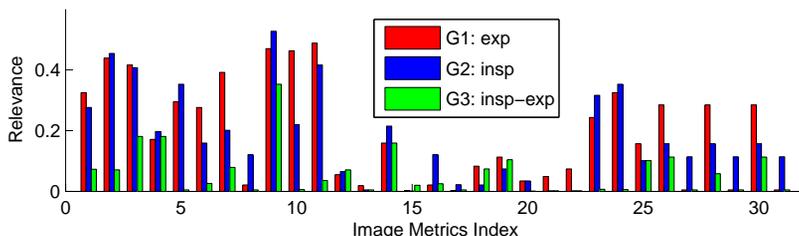


Fig. 2. Relevance (mutual information) of image metrics to different disease types. The x axis has the image metrics index listed in Table 2. Red bars are for metrics from $G1$, the expiratory images; blue bars for $G2$, the inspiratory images; and green bars for $G3$, difference of $G1$ to $G2$.

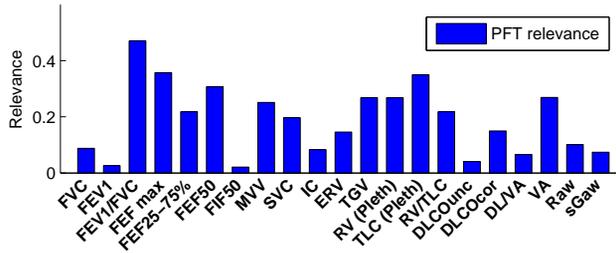


Fig. 3. Relevance (mutual information) of PFT values to different disease types. The x axis lists all the 21 PFT values in Table 1. Note that the range of the y axis is similar to the range from those image metrics in Fig. 2.

other words, a selected feature will provide additional information supporting the diagnosis that is missing from other features.

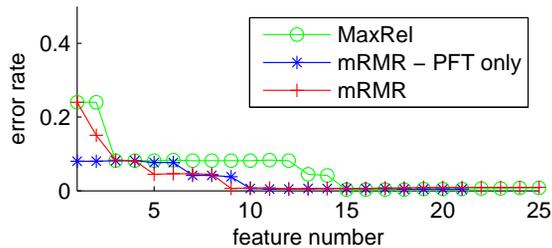
Rank	Feature	Rank	Feature
1	<i>G2</i> :5% attenuation value	14	<i>G3</i> :entropy
2	<i>G1</i> :kurtosis	15	<i>G1</i> :short run low grey level emphasis
3	VA	16	<i>G3</i> :sigma
4	<i>G1</i> :low grey level run emphasis	17	<i>G2</i> :mean
5	<i>G2</i> :run length nonuniformity	18	TLC (<i>pleth</i>)
6	<i>G1</i> :5% attenuation mean	19	<i>G2</i> :short run low grey level emphasis
7	FEF max	20	<i>G2</i> :sum
8	<i>G1</i> :95% attenuation value	21	<i>G2</i> :5% attenuation mean
9	<i>G1</i> :cluster prominence	22	<i>G3</i> :relative volume _{<-910}
10	<i>G1</i> :emphysema relative ratio	23	<i>G2</i> :relative volume _{<-910}
11	<i>G1</i> :volume	24	<i>G3</i> :cluster prominence
12	<i>G2</i> :grey level nonuniformity	25	<i>G1</i> :long run low grey level emphasis
13	FEV1/FVC		

Table 3. The first 25 selected features from mRMR [11]. Note these features include both PFT values and image metrics. The image metrics includes both the first order statistical measurements obtainable from the attenuation histogram and other more sophisticated metrics of texture descriptions.

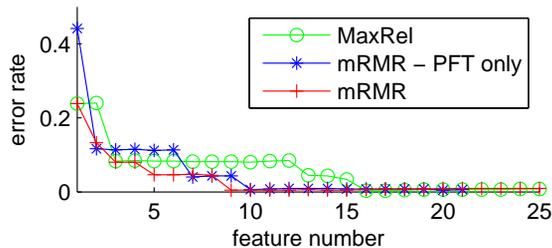
Table 3 lists all the 25 selected features using mRMR. Both PFT values and image metrics are selected. This indicates that image metrics provide extra information for differentiation of ILD and COPD, comparing with using only PFT values. Also image features computed both from the first-order statistical measurements (like the attenuation value) and from the texture descriptions (like the grey level run emphasis and the cluster prominence) are selected in the same subset. This also suggests that both types of image features are valuable for diagnosis. Out of these 25 features, 20 features statistically differ in ILD from

COPD groups using an unpaired t-test (null hypothesis rejected at 2% level), which means that statistical significance exists on our relatively small datasets.

In comparison to feature selection using mRMR, the baseline used in the paper is to rank features by their relevance values (i.e. $I(z_i, c)$ in Equ.1) and such a criterion is denoted as MaxRel as in [11]. In this paper, we use SVMs as classifiers to compare the efficiency of the first $n = 1$ to 25 features ranked by mRMR and MaxRel by increasing the size (n) of the selected feature set. In each test, one subject is excluded when training SVMs and that subject is used as the test set. The error rates are computed as the average ratio of wrongly classified subject by repeating using each of the patients as the test subject.



(a) SVM with Gaussian kernel



(b) SVM with linear kernel

Fig. 4. Classifying ILD and COPD using the first $n = 1$ to 25 selected features with (a) SVM with Gaussian kernel and (b) SVM with linear kernel, using MaxRel, mRMR and mRMR with only PFT values. For each curve, the x axis is the number of features; the y axis is the error rate. One subject is excluded when learning SVM parameters and that subject is used to test the accuracy of the learned SVM.

We test two SVM models with different kernels. The first is SVM with Gaussian kernel, which is a popular choice for general classification tasks [19]. Features are normalized by dividing them by the maximum the absolute values. To further investigate the efficiency of the selected features and isolate the influence of choice of kernels, we also test SVM with the most simple linear kernel. As shown in both Fig. 4 (a) and (b), mRMR (red curves with crosses) reaches low error rate using 9 features in comparison to MaxRel using 15 (SVM with Gaus-

sian kernel) or 16 (SVM with linear kernel). This shows the efficiency of mRMR framework.

To further demonstrate that image features are not redundant to the PFT values, we apply mRMR only on the subset of 21 PFT values and have a similar experiment by computing error rates with increasing selected feature set size. The performance is shown as “mRMR - PFT only” (the blue curves with stars) in Fig. 4. Although PFT values have similar error curves using SVM with Gaussian kernels (Fig. 4 (a)), it is inferior using linear kernels (Fig. 4). This can be explained as the selected image features are able to compensate the inefficiency of PFT features.

4 Conclusions

Various image metrics have been proposed and used in the literature to differentiate ILD and COPD. In this paper we studied their relevance values to the corresponding diagnoses in comparison with PFT values. Using a minimal-redundancy-maximal-relevance (mRMR) framework, we further looked into the redundancy between image metrics and PFT values. The result shows that not only some image metrics have similar relevance values compared with PFT, but also image metrics are not redundant when PFT values are provided. Image metrics of the first order statistics from the attenuation histogram and of more sophisticated texture descriptions are both selected, which again suggests that these two types of image metrics are both valuable for a further investigation. We finally gave a selected optimal subset of 25 features for characterization of ILD and COPD, including both image metrics and PFT values.

It should be noted that we did not go far enough in achieving an automated diagnosis of ILD and COPD system in this paper. Although certainly plausible in the future, such an automated diagnosis system, however, is premature at this point and not what we intended to develop in this paper. Rather, the focus of this paper is the demonstration that various image metrics are capable of providing significant information in characterizing clinical diagnosis compared to the gold standard of PFT values for ILD and COPD pathologies.

References

1. Miller, N.L., Staples, C.A., Miller, R.R., Abboud, R.T.: “density mask”. an objective method to quantitate emphysema using computed tomography. *Chest* **94**(4) (Oct 1988) 782–787
2. Gevenois, P.A., Vuyst, P.D., Sy, M., Scillia, P., Chaminade, L., de Maertelaer, V., Zanen, J., Yernault, J.C.: Pulmonary emphysema: quantitative CT during expiration. *Radiology* **199**(3) (Jun 1996) 825–829
3. Lamers, R.J., Kemerink, G.J., Drent, M., van Engelshoven, J.M.: Reproducibility of spirometrically controlled CT lung densitometry in a clinical setting. *Eur Respir J* **11**(4) (Apr 1998) 942–945

4. Xu, Y., Sonka, M., McLennan, G., Guo, J., Hoffman, E.A.: MDCT-based 3-D texture classification of emphysema and early smoking related lung pathologies. *IEEE Trans Med Imaging* **25**(4) (Apr 2006) 464–475
5. Zaporozhan, J., Ley, S., Eberhardt, R., Weinheimer, O., Iliyushenko, S., Herth, F., Kauczor, H.U.: Paired inspiratory/expiratory volumetric thin-slice CT scan for emphysema analysis: comparison of different quantitative evaluations and pulmonary function test. *Chest* **128**(5) (Nov 2005) 3212–3220
6. Coxson, H.O., Rogers, R.M., Whittall, K.P., D'yachkova, Y., Par, P.D., Sciruba, F.C., Hogg, J.C.: A quantification of the lung surface area in emphysema using computed tomography. *Am J Respir Crit Care Med* **159**(3) (Mar 1999) 851–856
7. Stavngaard, T., Shaker, S.B., Bach, K.S., Stoel, B.C., Dirksen, A.: Quantitative assessment of regional emphysema distribution in patients with chronic obstructive pulmonary disease (copd). *Acta Radiol* **47**(9) (Nov 2006) 914–921
8. Madani, A., Zanen, J., de Maertelaer, V., Geveno, P.A.: Pulmonary emphysema: objective quantification at multi-detector row CT—comparison with macroscopic and microscopic morphometry. *Radiology* **238**(3) (Mar 2006) 1036–1043
9. Temizoz, O., Etlik, O., Sakarya, M.E., Uzun, K., Arslan, H., Harman, M., Demir, M.K.: Detection and quantification of the parenchymal abnormalities in emphysema using pulmo-CT. *Comput Med Imaging Graph* **31**(7) (Oct 2007) 542–548
10. Vapnik, V.N.: *The nature of statistical learning theory*. Springer-Verlag New York, Inc. (1995)
11. Peng, H., Long, F., Ding, C.: Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans Pattern Anal Mach Intell* **27**(8) (August 2005) 1226–1238
12. Haralick, R.M., Shanmugam, K., Dinstein, I.: Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*. **3**(6) (1973) 610–621
13. Dasarathy, B.R., Holder, E.B.: Image characterizations based on joint gray-level run-length distributions. *Pattern Recognit. Lett.* **12** (1991) 497–502
14. Hu, S., Hoffman, E.A., Reinhardt, J.M.: Automatic lung segmentation for accurate quantitation of volumetric X-ray CT images. *IEEE Trans Med Imaging* **20**(6) (Jun 2001) 490–498
15. Hoffman, E.A., Simon, B.A., McLennan, G.: State of the art. a structural and functional assessment of the lung via multidetector-row computed tomography: phenotyping chronic obstructive pulmonary disease. *Proc Am Thorac Soc* **3**(6) (Aug 2006) 519–532
16. Stolk, J., Putter, H., Bakker, E.M., Shaker, S.B., Parr, D.G., Piitulainen, E., Russi, E.W., Grebski, E., Dirksen, A., Stockley, R.A., Reiber, J.H.C., Stoel, B.C.: Progression parameters for emphysema: a clinical investigation. *Respir Med* **101**(9) (Sep 2007) 1924–1930
17. Chu, A., Sehgal, C.M., Greenleaf, J.F.: Use of gray value distribution of run lengths for texture analysis. *Pattern Recognition Letters* **11** (1990) 415–420
18. <http://research.janelia.org/peng/proj/mRMR/>
19. Schölkopf, B., Smola, A.J.: *Learning with kernels : support vector machines, regularization, optimization, and beyond*. Adaptive computation and machine learning. MIT Press (2002)