Quantitative Assessment of Emphysema from Whole Lung CT Scans: Comparison with Visual Grading

Brad M. Keller[*a], Anthony P. Reeves[b], Tatiyana V. Apanasovich[c], Jianwei Wang[d], David F. Yankelevitz[d], Claudia I. Henschke[d]

[a]Department of Biomedical Engineering, Cornell University, Ithaca, NY, USA;
[b]School of Electrical and Computer Engineering, Cornell University, Ithaca, NY, USA;
[c]Department of Pharmacology and Experimental Therapeutics, Jefferson Medical College of Thomas Jefferson University, Philadelphia, PA, USA;
[d]Department of Radiology, Weill Medical College of Cornell University, New York, NY, USA;

ABSTRACT

Emphysema is a disease of the lungs that destroys the alveolar air sacs and induces long-term respiratory dysfunction. CT scans allow for imaging of the anatomical basis of emphysema and for visual assessment by radiologists of the extent present in the lungs. Several measures have been introduced for the quantification of the extent of disease directly from CT data in order to add to the qualitative assessments made by radiologists. In this paper we compare emphysema index, mean lung density, histogram percentiles, and the fractal dimension to visual grade in order to evaluate the predictability of radiologist visual scoring of emphysema from low-dose CT scans through quantitative scores, in order to determine which measures can be useful as surrogates for visual assessment. All measures were computed over nine divisions of the lung field (whole lung, individual lungs, and upper/middle/lower thirds of each lung) for each of 148 low-dose, whole lung scans. In addition, a visual grade of each section was also given by an expert radiologist. One-way ANOVA and multinomial logistic regression were used to determine the ability of the measures to predict visual grade from quantitative score. We found that all measures were able to distinguish between normal and severe grades ($p<0.01$), and between mild/moderate and all other grades ($p<0.05$). However, no measure was able to distinguish between mild and moderate cases. Approximately 65% prediction accuracy was achieved from using quantitative score to predict visual grade, with 73% if mild and moderate cases are considered as a single class.

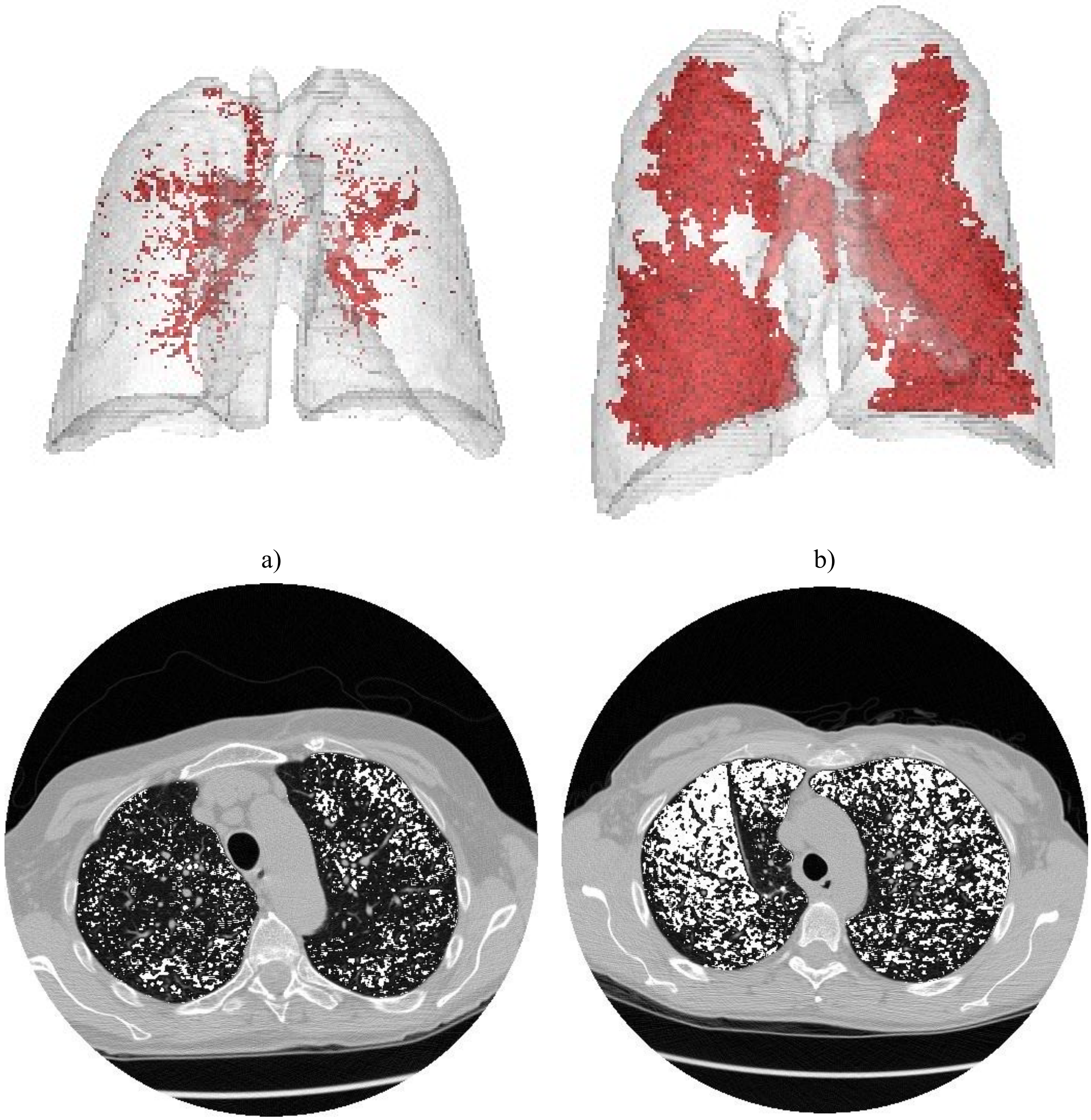
Keywords: CT, computer aided diagnosis, COPD, emphysema, lung disease, visual score

1. INTRODUCTION

High-resolution, multi-row detector computed tomography has allowed radiologists to view the underlying anatomical basis of chronic obstructive pulmonary diseases (COPD), particularly emphysema, from CT scans. Emphysema is clinically defined as the destruction and breakdown of the alveolar air sacs in the lung into large, coalescing pockets of anatomically dead-air space, emphysematous regions are visually described in CT images as being regions of lung parenchyma that are of a significantly low density. This has allowed for qualitative scoring of the disease state of an individual by radiologists. Computer-based scoring systems have been developed that extend this concept to allow for quantitative evaluation of emphysema from CT scans, with the majority of methods focusing on the use of density information as the primary index, either through relative area, such as the emphysema index, or distribution of regions, such as the fractal dimension. The emphysema index, also referred to as low attenuation area percentage (LAA%), developed by Müller [1], is the most well known of all these measures. Both visual and quantitative scoring have been showing to be relatively accurate in quantifying emphysema. [2]

Little work has gone into the evaluation of using quantitative measurements as a surrogate scoring system for visual scoring. The purpose of this work is to evaluate the predictability of radiologist visual scoring of emphysema, both globally and of sub-regions of the lung parenchyma, from low-dose CT scans through the analysis of four commonly used quantitative measurements. This work focuses on the four most commonly used scores reported in the literature: the emphysema index [3][4], the n-th percentile of the histogram [5][3], the mean lung density [6][7] and the fractal dimension [8][9].

Medical Imaging 2009: Computer-Aided Diagnosis, edited by Nico Karssemeijer, Maryellen L. Giger
Proc. of SPIE Vol. 7260, 726008 · © 2009 SPIE · CCC code: 1605-7422/09/$18 · doi: 10.1117/12.812254



a) b)

c) d)

Fig. 1. Visual representation of Emphysema Distribution in Whole Lung a) Mild Emphysema – 3D visualization b) Severe Emphysema – 3D visualization c) Mild Emphysema – CT slice d) Severe Emphysema – CT slice

2. METHOD

The introduction of high-resolution, multi-row detector CT has allowed radiologists to view the underlying anatomical basis of emphysema from CT scans. Given that emphysema is defined as the destruction and breakdown of the alveolar air sacs in the lung, emphysematous regions are visually described as being regions of lung parenchyma that are of a significantly low density. This allows for a qualitative scoring of the extent to which an individual has emphysema present in the lungs by radiologists investigating COPD and its effects. Computer-based scoring methods and systems have been developed that extend this concept to allow for quantitative evaluation of emphysema from CT scans. Emphysema index (EI) is the standard quantification measure of emphysema from lung CT-scans. Other commonly used and evaluated measures include mean lung density(MLD), histogram percentiles(HIST), and the fractal dimension(FD). All of these measures have been promoted as measures for quantification of the underlying anatomical basis for emphysema. However, little work to date has compared these four common measures to radiological visual scoring simultaneously. Therefore, this work focuses on the predictability of visual scores from quantitative measures. Figure 1 gives visual distributions of mild and severe emphysema in a whole lung scan.

Four primary measures were investigated in this work. The emphysema index is the classic measure of pulmonary emphysema and is the relative volume of the lung parenchyma that falls below a given density threshold. In this study, that threshold is set to -910 H.U. as that threshold is believed to cover. The emphysema index is also commonly called the percent low attenuation area (%LAA). The second measure of interest is the N^{th} percentile of the histogram. The histogram percentile is an empirical percentile, with the observation used in this case as the voxel densities of the lung, giving the density level at which N% of the lung pixel densities will fall below. Effectively the HIST measure returns the threshold needed to get an EI of N. In this work, N is set to be the 15^{th} percentile. The third measure is the fractal dimension. This returns a value that is indicative of the distribution of emphysema region sizes within the lung. The final measure is the mean lung density.

In order to determine the predictability of the visual grade from these measures, and thus which measures correlate most closely with radiologist observations, 148 low-dose, whole lung scans were randomly selected from a database of patients with known smoking history. Each scan was graded on the extent of emphysema present based on expert radiologist visual interpretation and was classified normal, mild, moderate, or severe. Nine classifications of the level of presence of emphysema in the lung was given by the radiologist for each of the 148 whole-lung scans. A single classification was given for the whole scan overall, for each of the two lungs independently, and for each of the three divisions of each lung independently (upper, middle, and lower third). The three classification schema can be considered as a global, lung, and six-partition schema. For each of the nine established divisions, the emphysema index and fractal dimension were computed at a -910 HU threshold using in-house software developed for use by radiologists in the Weill Medical College of Cornell University. The algorithms used were similar to the ones described by Müller et al. [1] and Mishima et al. [10] respectively. Mean lung density and the 15th percentile of the histogram were also computed using the same software.

One-way ANOVA using F-test and all pairwise comparisons among means are tested using Tukey-Kramer's HSD test were performed to show that means of objective scores associated with each visual score are significantly different from another. The equation for Tukey-Kramer's HSD is given below,

$$t_S = \frac{M_i - M_j}{\sqrt{\frac{MSE}{n_h}}}$$

where i and j represent the pair of means to be tested, $M_i - M_j$ is the difference between the i-th and j-th means, MSE is the Mean Square Error, and n_h is the harmonic mean of the sample sizes of groups i and j. The critical value of ts is determined from the distribution of the studentized range. The number of means in the experiment is used in the determination of the critical value, and this critical value is used for all comparisons among means. If t_s surpasses this critical value, the two means are considered significantly different. All statistical computations and tests in this work were performed with the use of the R statistical software package [11].

Table 1: Confusion Matrices of Classification of Emphysema Index to True Visual Grade in the six-partitioned lung segments Above) Mild and Moderate as separate classes Below) Mild and Moderate as a single class

EI \ Visual	Normal	Mild	Moderate	Severe
Normal	27	13	0	5
Mild	57	151	50	49
Moderate	0	0	0	0
Severe	13	74	64	372

EI \ Visual	Normal	Mild-moderate	Severe
Normal	16	9	1
Mild-moderate	73	247	89
Severe	8	96	337

Table. 2. Classification Accuracies for Measures and Lung Division Schema Left) Mild and Moderate as separate classes Right) Mild and Moderate as a single class

	Global	Lung	Regional
EI	64.2%	67.2%	62.3%
HIST	65.5%	67.2%	61.9%
MLD	59.5%	62.1%	58.4%
FD	66.2%	67.9%	59.5%

	Global	Lung	Regional
EI	73.0%	73.6%	68.5%
HIST	77.0%	75.7%	70.5%
MLD	67.6%	67.6%	64.4%
FD	75.0%	74.6%	64.2%

In order to determine the overall usefulness in using quantitative score to predict the subjective visual score for a given lung region, multinomial logistic regression was used to compute the probability that a objective score would predict a visual grade and classification accuracy was computed using each measure. Finally, a pair-wise implementation of the McNemar's test was used to determine if the four objective scores are significantly different in predicting visual score.

3. DATA

All scans used in this study were acquired at the Weill Medical College of Cornell University using a whole lung, low dose protocol. 148 total scans were acquired in this manner. 135 scans were at 1.25mm slice thickness acquired at 120 kVp; 13 scans were at 5mm slice thickness acquired at 140 kVp. All scans underwent the same visual grading procedure and quantitative scoring methods.

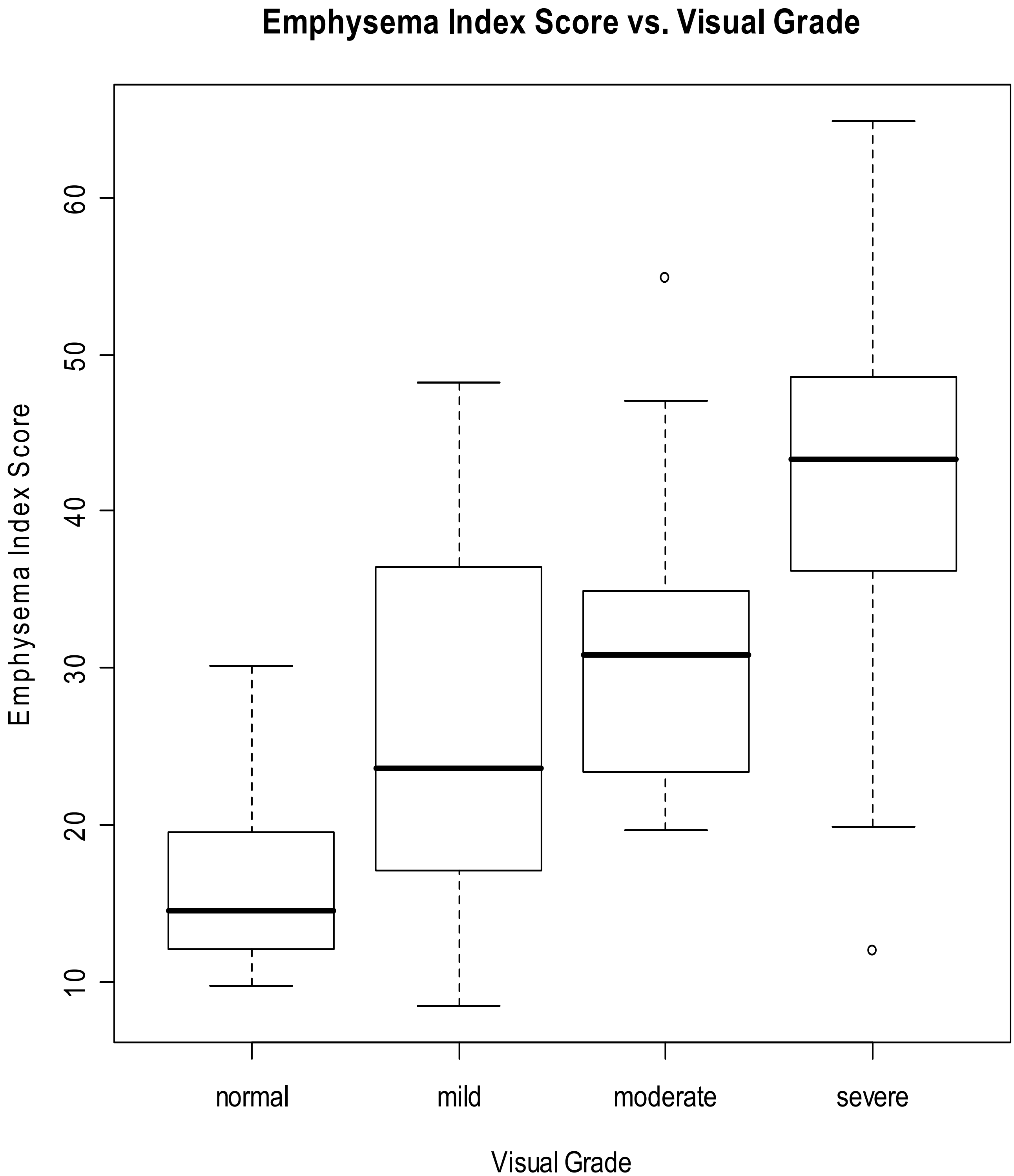


Fig. 2. Box plot of distributions of Emphysema Index vs. Visual Grade. Both normal and severe grades were shown to be statistically different from all groups ($p<0.05$) for all three lung partition schema. Of particular interest is the large overlap between mild and moderate emphysema scores.

4. RESULTS

We found that for all measures, both normal and severe grades were shown to be statistically different from all groups ($p<0.05$) for all three lung partition schema. However mild and moderate did not show statistical significance, and a large overlap in the ranges of scores for mild and moderate was found, with moderate cases tending to have quantitative scores in the upper range of mild cases. Figure 2 provides a box-plot that shows this effect for emphysema index in the six-partition schema. The average correct classification rate of visual score from quantitative measurements was approximately 65%, with greatest confusion occurring between mild and moderate grades due to the classification schema failing to separate mild and moderate cases, classifying all 'moderate' cases as 'mild' or 'severe'. Accuracies of over 75% could be achieved when mild and moderate were combined into single class to minimize this confusion. Table 1 illustrates this effect for emphysema index on the regional level of the lungs. Overall, FD, EI, and HIST were found to be superior to MLD at the global and lung levels ($p<0.1$) and EI and HIST being superior to FD and MLD at the regional level ($p<0.1$) for predicting visual grade from quantitative score. Table 2 gives the overall classification accuracies for the four measures across the three lung division analysis for both the four visual grade scoring method and three visual grade scoring method.

5. DISCUSSION

Little work has been done comparing multiple measures simultaneously to visual grading. Simultaneous comparison is important to show true advantages between various methods. This work shows that automated measures show good agreement with visual grade, particularly emphysema index, and these scores could be used as surrogate scores for the interpretation and assessment of emphysema severity in lung CT scans.

The findings in this work show that the quantitative scores can most easily distinguish between normal and severe cases of emphysema, regardless of the lung partition schema used. All four quantitative measures could also distinguish between normal and mild cases as well as between moderate and severe, although the distinction between these groups is not as clear as between normal and severe cases. However, no measure at any level was able to distinguish between mild and severe cases. The primary reason for this outcome is that in general, an average increase in emphysema score can be seen for the various measures when advancing visual grade from mild to moderate, the variation of scores falling within the mild case is enough to provide large overlap between scores of the mild and moderate cases. This is most clearly seen in figure 2, where approximately 50% of the moderate scores for emphysema index are shown to fall within a single quartile of the mild case. When the two categories are combined, we find that overall accuracies of classification increase for the measures by 10-15%, indicating that misclassification of moderate cases is the largest single source of error. Another possible reason is that there were fewer 'moderate' observations as opposed to 'mild' or 'severe', which, when combined with the reason above, could have ultimately biased the multinomial logistic regression to not classify observations as moderate.

When we compare the accuracy rates of three lung-division schema, a trend towards increased prediction accuracy of visual scores increases as we analyze larger sub-regions of the lung parenchyma at once, for example each whole lung individually versus analyzing the 3 divisions of a single lung, as is shown in Table 2. There are two possible explanations for this occurrence. The first is that the by analyzing divisions of a lung, the sample-size of the measurements taken increases substantially, which could account for increased accuracy in the larger. The second possible explanation is that the measures were originally designed to be used on whole lung scans, so that smaller variations within a sub-region could cause increased error rates, such as increased noise in the apices of the lungs. Normally, local variations would have a larger lung volume to be averaged against, thus mitigating their effects on the measures accuracy, but the limited regional volume could allow for the variations to have a larger effect. This could be considered a limitation of the measures for local analysis.

6. CONCLUSION

This work shows that there is a strong relationship between visual grade and quantitative measures when distinguishing between normal, mild/moderate, and severe cases. We also find that 65% predictability is possible for all divisions of the lung, with EI and HIST being the most indicative of visual grade. The prediction accuracy can be further increased to 75% if mild and moderate cases can be grouped as a single class. Given that automated measures show good agreement with visual grade, quantitative measures could be used as surrogate scores for the interpretation and assessment of emphysema severity in lung CT scans.

ACKNOWLEDGEMENTS

This research was supported in part by NIH grant R33CA101110 and the Flight Attendants' Medical Research Institute. Drs. Henschke, Yankelevitz and Reeves are co-inventors on a patent and other pending patents which are owned by Cornell Research Foundation (CRF) who non-exclusively licenses them to General Electric for technology involving detection and characterization of nodules in many medical situations, including measurement of nodules. They receive royalties from CRF pursuant to Cornell policy, which in turn is consistent with the Bayh-Dole Act. They are also inventors on a pending patent application for lesion measurement. Dr. Yankelevitz is an inventor on a pending patent owned by PneumRx related to biopsy needles, serves as a medical advisor to them, and holds an equity interest in PneumRx.

REFERENCES

[1] Muller, N.L., Staples, C.A., Miller, R.R. and R.T. Abboud, " 'Density Mask'. An Objective Method to Quantitate Emphysema Using Computed Tomography", Chest 94(4), 782-787 (1988).

[2] Park KJ, Bergin CJ, and Clausen JL, "Quantitation of Emphysema with Three-dimensional CT Densitometry: Comparison with Two-dimensional Analysis, Visual Emphysema Scores, and Pulmonary Function Test Results", Radiology 211, 541-547 (1999).

[3] Uppaluri, R., Mitsa, T., Sonka, M., Hoffman, E.A. and McLennan, G. "Quantification of Pulmonary Emphysema from Lung Computed Tomography Images", Am J Respir Crit Care Med 156(1), 248-254 (1997).

[4] Madani, A., de Maertelaer, V., Zanen, J. and P.A. Gevenois., "Pulmonary Emphysema: Radiation Dose and Section Thickness at Multidetector CT Quantification—Comparison with Macroscopic and Microscopic Morphometry", Radiology 243, 250-257 (2007).

[5] Gould, G.A., MacNee, W., McLean, A., Warren, P. M., Redpath, A., Best, J. J., Lamb, D. and Flenley, D. C., "CT Measurements of Lung Density in Life can Quantitate Distal Airspace Enlargement an Essential Defining Feature of Human Emphysema", Am Rev Respir Dis 137, 380-392 (1988).

[6] Best AC, Lynch AM, Bozic CM, Miller D, Grunwald GK, Lynch DA, "Quantitative CT Indexes in Idiopathic Pulmonary Fibrosis: Relationship with Physiologic Impairment", Radiology 228, 407-414 (2003).

[7] Ley S, Zaporozhan J, Morbach A, Eberle B, Gast KK, Heussel CP, Biedermann A, Mayer E, Schmiedeskamp J, Stepniak A, Schreiber WG, Kauczor HU, "Functional Evaluation of Emphysema Using Diffusion-Weighted 3Helium-Magnetic Resonance Imaging, High-Resolution Computed Tomography, and Lung Function Tests", Investigative Radiology.39(7), 427-434 (2004).

[8] Hoffman EA, Reinhardt JM, Sonka M, Simon BA, Guo J, Saba O, Chon Deokiee, Samrah S, Shikata H, Tschirren J, Palagyi K, Beck KC, McLennan G, "Characterization of the interstitial lung diseases via density-based and texture-based analysis of computed tomography images of lung structure and function", Academic Radiology 10(10), 1104-1118 (2003)

[9] Coxson HO, Whittall KP, Nakano Y, Rogers RM, Sciurba FC, Keenan RJ, Hogg JC, "Selection of patients for lung volume reduction surgery using a power law analysis of the computed tomographic scan", Thorax. 58(6), 510-514 (2003).

[10] Mishima, M., Hirai, T., Itoh, H., Nakano, Y., Sakai, H., Nishimura, K., Muro, S., Oku, Y., Chin, K., Ohi, M., Nakamura, T., Bates, J.H. T., Alencar, A. M. and Suki, B., "Complexity of Terminal Airspace Geometry Assessed

by Lung Computed Tomography in Normal Subjects and Patients with Chronic Obstructive Pulmonary Disease," Proc Natl Acad Sci 96(16), 8829-8834 (1999).
[11] R Development Core Team (2007). "R: A language and environment for statistical computing. R Foundation for Statistical Computing", Vienna, Austria. ISBN 3-900051-07-0, url: http://www.R-project.org.