Variation of Quantitative Emphysema Measurements from CT Scans

Brad M. Keller*[a], Anthony P. Reeves[b], Claudia I. Henschke[c],
R. Graham Barr[d], David F. Yankelevitz[c]
[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 Radiology, Weill Medical College of Cornell University, New York, NY, USA;
[d]Department of Medicine, College of Physicians and Surgeons, Columbia University, New York, NY, USA

ABSTRACT

Emphysema is a lung disease characterized by destruction of the alveolar air sacs and is associated with long-term respiratory dysfunction. CT scans allow for imaging of the anatomical basis of emphysema, and several measures have been introduced for the quantification of the extent of disease. In this paper we compare these measures for repeatability over time. The measures of interest in this study are emphysema index, mean lung density, histogram percentile, and the fractal dimension. To allow for direct comparisons, the measures were normalized to a 0-100 scale. These measures have been computed for a set of 2,027 scan pairs in which the mean interval between scans was 1.15 years (σ: 93 days). These independent pairs were considered with respect to three different scanning conditions (a) 223 pairs where both were scanned with a 5 mm slice thickness protocol, (b) 695 with the first scanned with the 5 mm protocol and the second with a 1.25 mm protocol, and (c) 1109 pairs scanned both times using a 1.25 mm protocol. We found that average normalized emphysema index and histogram percentiles scores increased by 5.9 and 11 points respectively, while the fractal dimension showed stability with a mean difference of 1.2. We also found, a 7 point bias introduced for emphysema index under condition (b), and that the fractal dimension measure is least affected by scanner parameter changes.

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

1. INTRODUCTION

The introduction of high-resolution, multi-row detector CT has allowed radiologists to view the 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. 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 or distribution of regions. The emphysema index, developed by Müller[1], is the most well known of all these measures and is illustrated in Figure 1. This work focuses on the four most commonly used scores reported in the literature: the emphysema index, the n-th percentile of the histogram, the mean lung density and the fractal dimension.

Ongoing CT lung cancer screening trials provide an opportunity to retro-actively look at population distributions of emphysema scores, as smoking is the primary risk factor for both lung cancer and emphysema[2]. Recently, there has been concern that the variation of these measures over time, due to varying inspiration levels[3] and altered scanner settings[3,4], would limit the usefulness in measuring disease progression. It is therefore important to quantitatively measure the distribution of these measures and quantify the biases introduced by varying scanner parameters, such as slice thickness.

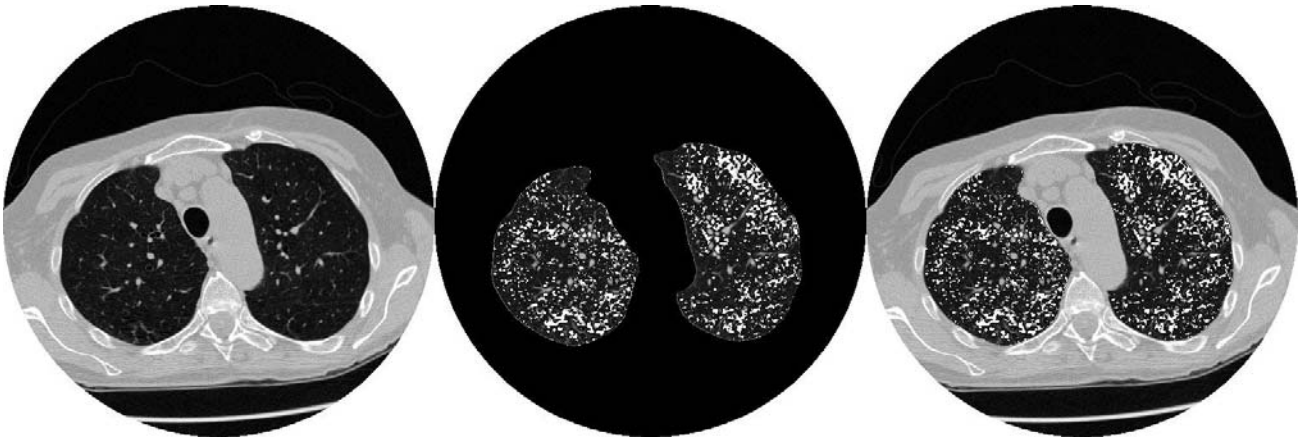


Figure 1: Sample emphysema index computed at -910 HU from a whole lung CT scan. Left: Standard axial CT slice. Center: White regions extracted denoting emphysematous regions defined by attenuation area of less than -910 HU. Right: Overlap of emphysema regions on original scan.

The aim of this study was to evaluate established quantitative emphysema measures from low-dose, whole-lung CT scans and determine measure distribution in two given scan acquisition protocols. The primary measures of interest are those most commonly used to quantify emphysema levels from CT scans: the emphysema index, mean lung density, histogram percentiles, and fractal dimension. All of these measures have been promoted as methods for accurate quantification of the underlying anatomical basis for emphysema. However, a major concern with the use of these emphysema measures has been their clinical lack of repeatability over longitudinal scans[2]. Therefore, the variation of these standard measures over approximately one year was also evaluated in the context of varying scan acquisition parameters.

2. METHODOLOGY

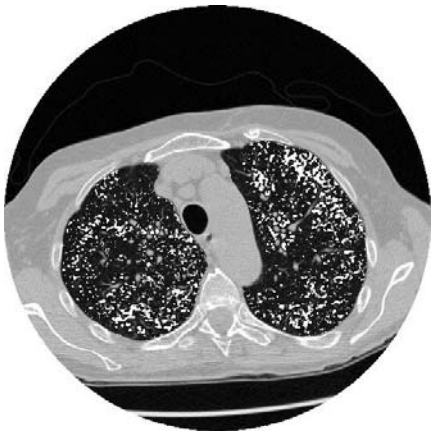
2.1 Quantitative emphysema measure acquisition

Four primary methods of emphysema level quantification were investigated in this work. The first, emphysema index[1] (EI), is the classic measure of pulmonary emphysema and is the relative volume of the lung that falls below a given density threshold. The emphysema index is also commonly called the percent low attenuation area (%LAA). In this study, the threshold used to evaluate is set to -910 H.U, as this level tends to count all voxels of even mild emphysema suspicion as emphysematous. As the emphysema index increases, more emphysema is present in the lungs. The second measure of interest is the N^{th} percentile of the histogram[5] (HIST). This returns the density value that would return an emphysema index of N. In this work, N is set to be the 15^{th} percentile, as that is most similar to the -910 threshold. A reduction in the n-th percentile density indicates a higher overall level of emphysema present in the lungs. The third measure is the mean lung density[6] (MLD), which is inversely proportional to the level. As overall density of the lung decreases, there is a rising level of emphysema present in the lungs. The final measure is the fractal dimension[7] (FD), as introduced by Mishima et al. The fractal dimension returns a value that is indicative of the distribution of emphysema region sizes versus the total number of regions within the lung. A common way of reporting the fractal dimension, given that the value is always negative, is to report the measure as 'alpha', which is the absolute value of the fractal dimension. As the value of alpha decreases, emphysema severity increases, as many small emphysema regions coalesce into fewer, larger regions.

For every scan analyzed in this work, 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. and Mishima et al. respectively. Mean lung density and the 15^{th} percentile of the histogram were also computed using the same software.

Measure	Normalized Score
EI_{-910}	13.9
HIST15	7
MLD	18.8
FD_{-910}	14.4

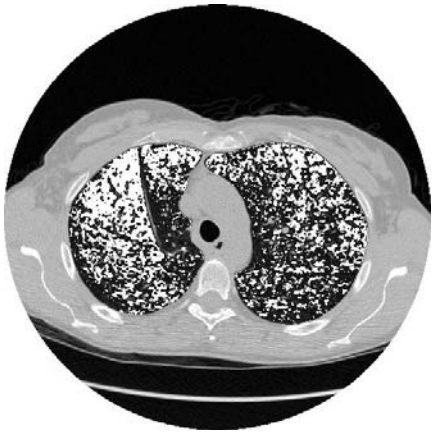
(a)



(b)

Measure	Normalized Score
EI_{-910}	44.5
HIST(15)	54
MLD	69
FD_{-910}	65

(c)



(d)

Figure 2: Emphysema measurement rescaling. Top: a) Normalized scores and b) sample slice for a mild case of emphysema. Bottom: c) Normalized scores and d) sample slice for a moderate-severe case of emphysema.

One issue that arises when comparing these measures to one another is that the measures investigated in this study are on different scales relative to one another. In essence, the emphysema index is a dimensionless percentage from 0-100, the mean lung density and 15th percentile of the histogram are in Houndsfield units on the order of less than -700 H.U, and the fractal dimension is on a smaller scale of approximately 0.5 to 3. In order to correctly compare these various measures, we therefore have to normalize the measures to be on the same scale as the emphysema index in order to bring them into agreement. This was done by empirically determining the effective range of the measures as seen in the dataset, and setting the to a 0 to 100 point scale. This was accomplished by calculating the values for the least and most severe levels of emphysema present for each of the measures within our dataset, and performing a linear transformation to convert that range to be 0-100, with 0 indicating no/negligible emphysema present and 100 indicating the most severe levels commonly seen in clinical practice.

Figure 2 demonstrates the normalization technique as done on a mild case (top row) and moderate to severe case of emphysema (middle row). The tables on the left give the normalized scores of the four measures investigated in this work for a given case, and right image is a sample slice with computed emphysema index, similar to Figure 1. The bottom graph gives a distribution of the normalized scores present in the study for the thick slice protocol data set described below.

2.2 Emphysema measurement distribution

The image data was taken when there was a change (upgrade) in CT scanner equipment, which is a standard occurrence in long term programs. Subsequently, the CT scanner protocol used in the study was altered in order to take advantage of the improved capabilities of the new equipment. This however leads to biases introduced for each measure on subsequent scans. In order to accurately compare longitudinal scans for a given patient, as well as accurately compare the distribution of measured diseased state for a population, these biases need to be quantified in order to correct for them.

In order to understand how the quantitative emphysema measures vary for a particular population, we used two datasets comprised of 2 different acquisition parameters to investigate the bias introduced when altering scan acquisition settings. The datasets are described in more detail in section 3. For each dataset, we computed the distribution of the emphysema measures, including the mean and empirical 90% confidence interval, for all four measures of interest in this study. For both of these datasets, we report the normalized measures, as this allows for direct comparison of the measures. We then compare the measures directly to one another across both parameter settings and look at both the change in mean and variation.

2.3 Emphysema measurement sensitivity to acquisition settings

To determine if there were differences in the inter-scan variation between these measures based on varying scanner parameters, the second aspect of this research looked at the inter-scan variability of the measures of interest. The available cases were sub-divided into 3 sets of scan-pairs by combining two sequential scans from a single case where the scanner parameters met specific criterion in the baseline scan and the sequential scan. The first two sets were comprised of scans that did not change scanner parameters in going from the initial to follow-up scans, that is that scanner type, slice thickness and dose were identical between scans in a given dataset. The first set used a thick slice protocol, and the second set used a thin slice protocol. The final set was comprised of scan-pairs where the first scan was acquired at the original scan parameter setting using the thick slice protocol and the follow-up scan was acquired with the upgraded scanners and the updated thin slice protocol. This is discussed in more detail in section 3.

For each of these three datasets, the inter-scan differences and absolute differences in value between each normalized measure was computed. In order to understand the variability of the measures overall, the median difference for absolute change as well as the empirical 90% confidence interval were calculated. Further comparing the three scan-pair datasets and their variation would allow for the identification and quantification of a bias introduced when changing protocol in a longitudinal study and possibly allow for a methodology to correct said bias.

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. To measure the bias introduced to quantitative emphysema measures as an indicator of disease state for a given population, two data sets representing CT scans taken using each of two parameter settings were acquired. The first dataset consisted of scans taken at a 5mm slice thickness at 140 kVP using a GE LightSpeed QX/i. There were 241 scans at this resolution. This parameter setting is referred to in this paper as parameter setting ‘A’. Subsequently, a second dataset consisting of 1293 scans was acquired using a different scanner using a 1.25mm slice thickness at 120 kVp using a GE LightSpeed Ultra. This parameter setting is referred to in this paper as parameter setting ‘B’.

In order to quantify the biases introduced by varying scanner type and acquisition parameters over a longitudinal study, three additional datasets of scan-pairs were acquired and analyzed. The first two sets were comprised of scans that did not change scanner parameters in going from the initial to follow-up scan. These were termed the 'A-to-A' set, as in first scan was acquired using parameter set A as was the follow-up scan, of which there were 223 pairings with a mean time difference between scans of 402 days, and the 'B-to-B' set, of which there were 1109 number of pairings with a mean time difference between scans of 414 days. A third dataset was compiled using 695 sequential scans that went from using parameter set A in the initial scan to parameter set B in the sequential scan. This set of pairings was labeled the 'A-to-B' set and had a mean time difference between scans of 439 days.

Table 1: Distribution of normalized measures for specific parameter settings. Reported values are mean and standard deviation (μ,σ)

Measure / Parameter	EI (-910)	HIST(15)	MLD	FD (-910)
Setting A	33.2, 15.0	41.6, 20.1	57.1, 13.7	50.9, 18.1
Setting B	39.1, 14.7	52.6, 32.4	54.0, 18.5	49.7, 20.1

Table 2: Absolute Differences between subsequent scans for normalized measures. Reported values are median and empirical 90% confidence interval

Measure / Dataset	EI_{-910}	HIST(15)	MLD	FD_{-910}
A-to-A	5.0 (0.4-19.7)	7 (0.1-26)	4.7 (0.7-19.0)	3.8 (0.3-17.1)
A-to-B	9.6 (1.1-25.2)	20 (4-48)	6.4 (0.6-27.5)	5.4 (0.5-26.5)
B-to-B	4.6 (0.3-16.0)	6 (1-25)	4.5 (0.5-21.0)	4.7 (0.3-19.6)

Table 3: Differences between subsequent scans for normalized measures. Reported values are mean and standard deviation (μ,σ)

Measure / Dataset	EI_{-910}	HIST(15)	MLD	FD_{-910}
A-to-A	3.6, 10.5	5.5,16.2	2.9, 11.3	-2.2, 9.9
A-to-B	8.8, 9.6	19.9,16.4	-4.4, 12.4	0.6, 12.6
B-to-B	-0.1, 8.3	-0.7, 13.1	0.7, 11.1	0.8, 10.3

4. RESULTS

Table 1 gives the mean and standard deviations for the distribution of the normalized measures for scans using the thick slice protocol, parameter setting 'A', and scans using the thin slice protocol, parameter setting 'B'. Reported values are mean and standard deviation. Figure 3 provides a visualization of these distributions, showing a graph of the histogram for each measure with a bin size of ten points.

Table 2 shows the median and empirical 90% confidence interval of absolute differences between normalized measures in subsequent scans. 'A-to-A' denotes the first data set where both the initial scan and follow-up scan are acquired using parameter setting A as previously described. 'A-to-B' denotes the data set where the initial scan is taken using parameter setting A and the follow-up scan is acquired using parameter setting B. 'B-to-B' denotes both scans being acquired using parameter setting B. When comparing constant inter-scan parameter settings (A-to-A and B-to-B) to altered settings (A-to-B), we found the median absolute variation to increase between 25% for the fractal dimension to approximately 250% for the histogram percentile. The range of variation also increased in a similar manner for the A-to-B dataset versus the constant datasets.

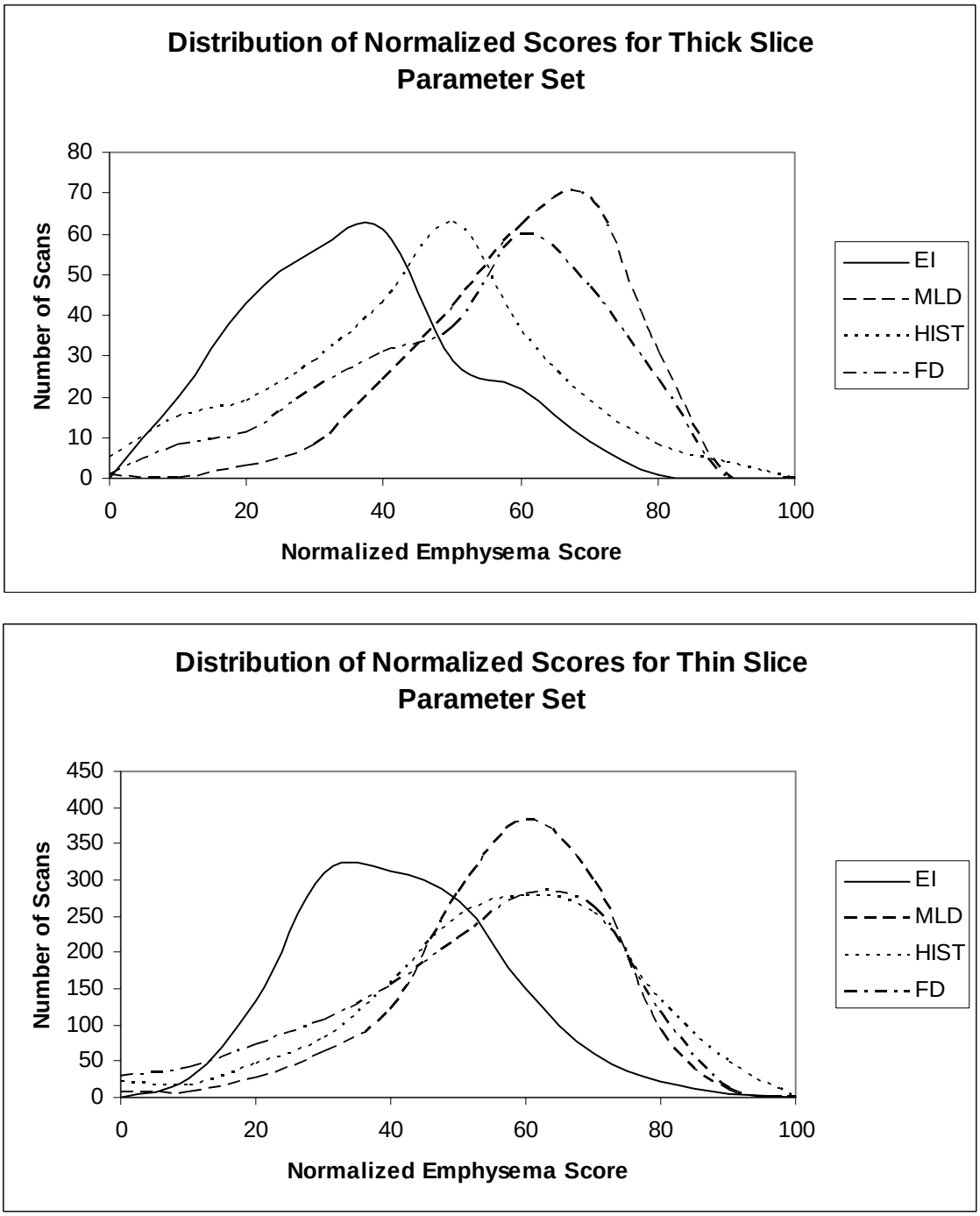


Figure 3: Distribution profiles of normalized emphysema scores for two protocols. Curves were fit to each histogram with a bin size of 10 points. Top: Distribution in thick slice (5mm) protocol. Bottom: Distribution in thin slice (1.25mm) protocol.

Table 3 shows the mean and standard deviation of the inter-scan difference of the normalized measures for each of the three parameter setting combinations, as described for Table 2. When comparing the A-to-B cohort to the consistent parameter pairings, there was a trend towards a bias for all measures except the fractal dimension.

5. DISCUSSION

Comparing the differences in normalized emphysema scores between longitudinal scans for both constant and varying scan acquisition parameters allows for the determination of biases introduced by the specific protocol change. We found that the mean difference was greater for the varying parameter dataset for all the measures other than the fractal dimension. The histogram percentile score was found to be the most affected by the change in scan acquisition protocol. This would imply that the fractal dimension could be very useful as a of measure of emphysema in long-term studies as it appears to be the least sensitive to the variation and biases introduced by changes in scan acquisition protocol.

The inter-scan variability of quantitative measures of emphysema is important to understand in longitudinal studies in order to accurately asses true change in a person's status versus random variability in the measure. Previous studies have found that CT scanner settings and dosing[8], as well as the level of inspiration[3], can have significant effects on quantitative measurements. However, no study to the author's knowledge has compared several of these quantitative measures concurrently and on a very large cohort. In that context, Table 2 shows that while emphysema index, mean lung density and the fractal dimension had comparable median differences, with the histogram percentile tending to have the largest variation. We also found a slight trend toward emphysema index and fractal dimension to have similar variation and that both are superior when compared to the other measures.

Inter-scan alteration of scanner parameters was also found to impact the overall variability of the emphysema measures. When comparing scan-pairs where both scans are acquired using the same settings to scan-pairs acquired with altered settings, we found an large increase in measure variability across the various measures. Overall the fractal dimension had the least amount of change in variability due to varying scanner parameter settings. This indicates that the fractal dimension is the most robust of the four measures against the biases introduced by altered scan acquisition protocols.

From table 1, we see that on the linear scale used for normalization that mean lung density, histogram percentile and fractal dimension all give higher scores on average versus the emphysema index. Figure 3 allows for further visualization.

6. CONCLUSION

We computed the four most common emphysema measures from CT scans on a dataset of 2027 scan pairs. We found that the fractal dimension was least affected by changes in scan acquisition protocol when compared to the emphysema index, histogram percentile, and mean lung density. We also found that emphysema index and fractal dimension have similar levels of variation to one another and are superior to the other measures in this regard.

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] Gietema, H.A., Schilham, A.M., van Ginneken, B., van Klaveren,, R.J., Lammers, J.W.J. and Prokop, M., "Monitoring of Smoking-induced Emphysema with CT in a Lung Cancer Screening Setting: Detection of Real Increase in Extent of Emphysema", Radiology 244(3), 890-897 (2007).
[3] Stoel, B.C. and Stolk, J., "Optimization and Standardization of Lung Densitometry in the Assessment of Pulmonary Emphysema," Investigative Radiology 39(11), 681-688 (2004).
[4] Boedeker, K.L., McNitt-Gray, M.F., Rogers, S.R., Truong, D.A., Brown, M.S., Gjertson, D.W. and Goldin, J.G., "Emphysema: Effect of Reconstruction Algorithm on CT Imaging Measures," Radiology 232, 295-301 (2004).

[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] 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).
[7] 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).
[8] 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).