Published in final edited form as:
J Thorac Imaging. 2013 September ; 28(5): . doi:10.1097/RTI.0b013e3182a21969.

Quantitative CT Imaging of Interstitial Lung Diseases

Brian J Bartholmai, M.D.,
Department of Radiology, Mayo Clinic, 200 First St SW, Rochester, MN 55905

Sushravya Raghunath, MS.,
Department of Physiology and Biomedical Engineering, Mayo Clinic, 200 First St SW, Rochester, MN 55905

Ronald A Karwoski,
Department of Physiology and Biomedical Engineering, Mayo Clinic, 200 First St SW, Rochester, MN 55905

Teng Moua, M.D.,
Division of Pulmonary and Critical Care Medicine, Mayo Clinic, 200 First St SW, Rochester, MN 55905

Srinivasan Rajagopalan, Ph.D.,
Department of Physiology and Biomedical Engineering, Mayo Clinic, 200 First St SW, Rochester, MN 55905

Fabien Maldonado, M.D.,
Division of Pulmonary and Critical Care Medicine, Mayo Clinic, 200 First St SW, Rochester, MN 55905

Paul A Decker, and
Division of Biomedical Statistics and Informatics, Mayo Clinic, 200 First St SW, Rochester, MN 55905

Richard A Robb, Ph.D.
Department of Physiology and Biomedical Engineering, Mayo Clinic, 200 First St SW, Rochester, MN 55905

Brian J Bartholmai: Bartholmai.Brian@Mayo.edu; Sushravya Raghunath: Raghunath.Sushravya@Mayo.edu; Ronald A Karwoski: Karwoski.Ronald@Mayo.edu; Teng Moua: Moua.Teng@Mayo.edu; Srinivasan Rajagopalan: Rajagopalan.Srinivasan@Mayo.edu; Fabien Maldonado: Maldonado.Fabien@Mayo.edu; Paul A Decker: Decker.Paul@Mayo.edu; Richard A Robb: Robb.Richard@Mayo.edu

Abstract

Purpose—High-Resolution chest CT (HRCT) is essential in the characterization of interstitial lung disease (ILD). The HRCT features of some diseases can be diagnostic. Longitudinal monitoring with HRCT can assess progression of ILD; however, subtle changes in the volume and character of abnormalities can be difficult to assess. Accuracy of diagnosis can be dependent on expertise and experience of the radiologist, pathologist or clinician. Quantitative analysis of thoracic HRCT has the potential to determine the extent of disease reproducibly, classify the types of abnormalities and automate the diagnostic process.

Corresponding Author: Brian J Bartholmai, MD, Department of Radiology, Mayo Clinic, 200 First St SW, Rochester, MN 55905, Phone: 507-284-4292, FAX: 507-284-8996, Bartholmai.Brian@Mayo.edu.

Materials and Methods—Novel software that utilizes histogram signatures to characterize pulmonary parenchyma was used to interrogate chest HRCT data, including retrospective processing of clinical CT scans and research data from the Lung Tissue Research Consortium (LTRC). Additional information including physiologic, pathologic and semi-quantitative radiologist assessment was available to allowcomparison of quantitative results with visual estimates of disease, physiologic parameters and measures of disease outcome.

Results—Quantitative analysis results were provided in regional volumetric quantities for statistical analysis as well as a graphical representation. Analysis suggests that quantitative HRCT analysis can serve as a biomarker with physiologic, pathologic and prognostic significance.

Conclusion—It is likely that quantitative analysis of HRCT can be used in clinical practice as a means to aid in identifying probable diagnosis, stratifying prognosis in early disease, and consistently determining progression of disease or response to therapy. Further optimization of quantitative techniques and longitudinal analysis of well-characterized subjects would be helpful to validate these methods.

Keywords

interstitial lung disease; CT; quantitative analysis; imaging biomarkers; computer aided detection

Introduction

Lung diseases, including interstitial lung diseases (ILD), remain a major cause of morbidity and mortality[1]. Radiologic evaluation through high-resolution CT (HRCT) has become increasingly essential to the characterization and classification of ILD. The characteristic imaging findings can be diagnostic for some pathologic processes such as idiopathic pulmonary fibrosis/usual interstitial pneumonitis (IPF/UIP). The ability of radiologic evaluation to differentiate between diseases such as UIP, that have an associated mortality approaching 75% five years after the diagnosis, and other diseases with less ominous prognosis has played a key role in patient management[2]. Other ILDs such as hypersensitivity pneumonitis, nonspecific interstitial pneumonitis, acute interstitial pneumonitis, desquamative interstitial pneumonitis, respiratory bronchiolitis-associated interstitial lung disease, lymphoid interstitial pneumonia and cryptogenic organizing pneumonitis can all exhibit differentiating features on HRCT. However, these distinctive disease processes often have a very similar clinical phenotype, and can also have indeterminate pathologic and radiographic appearances. In addition, some patients can have mixed restrictive/fibrotic and destructive/obstructive processes, such as in combined pulmonary fibrosis and emphysema syndrome[3], which confound physiologic testing and for which biopsy results can vary wildly based on the location of sampling. The fundamental clinical problems of how to consistently detect, characterize, and differentiate the various ILDs remain diagnostic challenges. Yet, all of the various ILD and mixed parenchymal diseases have distinctly different prognoses and opportunities for therapy[4] and it is increasingly clear that specific therapy targeted on a particular pathological process will be key to altering the progression of these generally inexorable diseases. Furthermore, how even known processes can be consistently characterized and quantified over time and, more importantly, how these changes can predict prognosis all remain largely unanswered questions. The promise of early and appropriate application of therapy so as to significantly affect prognosis has not been achieved despite many efforts, and assessment of response to therapy has been difficult in even well-designed, randomized prospective trials – in part due to lack of robust non-invasive biomarkers for ILD.

The general rationale for developing a method for quantitative analysis of the lung parenchyma derives from the knowledge that the generic term ‘ILD’ includes multiple

different diseases with numerous different imaging features in variable distribution. In many cases, the patterns and distribution of disease may allow differentiation of these processes. Specifically, HRCT is able to demonstrate features of disease, with visual patterns such ground glass opacities, reticular infiltrates and honeycombing that enable the differential diagnosis to be narrowed[5]. The impact of HRCT on clinical management is reflected in the decreased need for surgical lung biopsy between the diagnostic workflow proposed in the ATS/ERS consensus statement in 2002[6] and more recent evidence-based guidelines for diagnosis and management in clinical practice[7,8]. Even in cases where the HRCT images lack the specific features that reflect the cellular infiltration, fibrosis and architectural distortion with honeycombing typical of a process such as UIP, HRCT remains a useful non-invasive technique to reveal the abnormal parenchymal densities resulting from microscopic morphological changes[9–15] and can provide guidance for the optimal site to obtain a biopsy of characteristic or active disease[12]. Similarly, it is generally accepted that the extent of visual abnormalities correlates with extent of pathological involvement as well as the severity of physiologic abnormalities[16], and therefore longitudinal HRCT can be useful in the monitoring of disease progression and response to therapy[17–19].

The complex morphological patterns of ILD that can change in extent and/or appearance over time can be challenging to assess. Similarly, manual classification and evaluation of extent is tedious and not reproducible. Evaluation is complicated by significant in inter- and intra-observer variation in the diagnosis of ILD[20,21]. Clearly, there are differences in the skill and experience of physicians involved in the process of determining a diagnosis, but there are more complex inherent differences in perception, interpretation of visual features of disease and 'reader error' which lead to variable description of findings and characterization of disease that may not be surmountable by any degree of training or advances in imaging technology[22–24]. Specifically, the final clinical diagnosis of ILD has been shown to be variable based on the training and experience of the radiologist, clinician and pathologist working independently or involved in the multidisciplinary evaluation of disease despite the presence of accepted criteria and clinical pathways even in a well-controlled investigation with well-characterized subjects. These variations are even more prevalent in real-world patient care within- and across-specialty physicians from academic and community health centers[25]. Even diagnosis by consensus of multiple experts and utilization of continuous learning techniques, which do improve consistency to some degree, does not assure dependable results[26].

The variability in clinical evaluation of ILD is an opportunity for automation, computer-aided detection and quantitative image analysis. Specifically, there are opportunities to use current image processing technology on CT data to optimize detection of the abnormalities recognized to represent pathological changes on HRCT and enable reproducible quantification and characterization of the manifestations of ILD. These quantitative results have promise as biomarkers and could result in more consistent diagnosis, more sensitive disease monitoring and accurate determination of prognosis[27]. There have been many different efforts using different methods to robustly characterize the discernible patterns in chest x-rays and CT scans over the last several decades but the assessment of results from some early texture-based analysis remains generally unchanged: "…*classification and quantification of interstitial lung disease is difficult, and even experienced chest radiologists frequently struggle with differential diagnosis. Automated schemes that indicate a percentage of affected lung or the probability of a certain disease would certainly be welcome, but require much more research*"[28]. Encouragingly, quantitative methods used to assess the severity of emphysema and other features of COPD on CT have become more robust over the last 20 years, with evolution and optimization of various techniques. These quantitative analysis results can play a significant role as biomarkers used in the diagnosis of COPD phenotypes, assessment of disease progression and prognosis[29]. However, since ILD

changes on HRCT are even more complicated than the density changes seen with emphysema, the development of quantitative CT-based measures for lung fibrosis has been more challenging[27–30] and results less encouraging. In general, simple methods based on pixel counting, first order features based on density masks or whole-lung histogram analysis, methods using multiple higher-order features or texture methods, and even more sophisticated classification techniques including continuous learning with physician-in-the-loop are only partially successful for evaluation of specific diseases or even simple determination of normal vs. abnormal regions[30–32]. Furthermore, many of the current methods used in evaluation of ILD for research purposes are computationally intense and require processing times that may take hours or even days. These real-world limitations make those techniques difficult to translate into routine clinical practice.

The purpose of the effort to develop software for analysis of thoracic HRCT is to provide a reproducible way to quantify and characterize the extent of diffuse parenchymal disease. We hypothesize that the results of parenchymal analysis can be correlated with radiologist visual assessment, accepted physiologic features of pulmonary disease, and clinical outcomes. The quantity, type and distribution of abnormalities by an automated tool should provide utility in clinical practice by aiding in non-invasive diagnostic determination, detecting change in disease over time, and stratifying risk of progression or mortality.

Methods

Towards addressing the need for improved quantitative analysis tools that can be utilized for both research analysis and clinical practice, a multidisciplinary team of Mayo Clinic clinicians, scientists and engineers at the Mayo Biomedical Imaging Resource have developed CALIPER (Computer Aided Lung Informatics for Pathology Evaluation and Rating): a computational platform for the near-real-time characterization and quantification of lung parenchymal patterns on CT scans. The following briefly outlines the computational components and methods utilized to test the efficacy of CALIPER.

Initial handling of data within CALIPER involves segmentation and extraction of anatomic regions as a pre-processing step towards the eventual characterization and quantification of the pulmonary parenchyma. The segmentation of the lungs in datasets was achieved using an adaptive density-based morphology. Lung extraction was performed using optimal thresholding to identify low density fields in the scans, region growing and void filling[33]. Three-dimensional connected low density components that touch the edge of the volume were assumed to be air outside of the thorax and eliminated from the volume of analysis. The lungs were isolated through connected component analysis and three dimensional hole filling was used to fill the lung cavities created by the elimination of normal blood vessels during the thresholding process. In the event the left and right lung continue to be connected after the above steps, layers of edge voxels were removed from the segmented lungs using morphological erosion iteratively, so as to break the connectivity of the lungs. When the eroded structures are separate, the removed edge voxels were re-assigned using conditional dilation. The lung borders were subjected to a final smoothing step using morphological closing with a 23×23×5 elliptical structuring element so as to close the gaps in the lung volumes along the mediastinum created by pulmonary arteries and veins.

The airways were automatically segmented by iterative application of increasingly restrictive constraints to a thresholding and 3D region growing process. The goal of the iterative process is to segment the most complete tracheobronchial tree while excluding other low density lung regions such as emphysema, honeycombing or other gas-filled structures such as the stomach or colon. Airway extraction was first attempted utilizing the 1×1×3 neighborhood grey scale minimum, with a threshold of −950 HU and 26 neighbor

connectivity. Each axial slice of the result was then checked for the number of 2D connected components. If the number of connected components exceeded 50, the segmentation was assumed to include regions extraneous to airways and was rejected. In such cases, the segmentation was repeated using more restrictive parameters such as 6-neighborhood connectivity and −960 HU thresholding.

Pulmonary vessels were segmented using an optimized multi-scale tubular structure enhancement filter based on the eigenvalues of the Hessian matrix[34]. These filters calculate 2nd-order derivatives within the neighborhood of each of the lung voxel. The eigenvalues of the Hessian matrix constructed from the derivatives were analyzed to determine the likelihood of the underlying voxel belonging to a dense tubular structure and hence a vessel.

Given that there are visually distinct normal anatomic and disease-specific morphological manifestations apparent on HRCT scans, parenchymal classification is typically approached though texture analysis, computer vision-based image understanding of volumetric histogram features and 3D morphology of the classified voxels. In general, the classifier results are determined by methods used to solve content-based information retrieval problems[28,35–40]. Central to all these schemes is the selection of representative expert-labeled volumes of interest (VOIs) as a training set for a classifier. The features of these VOIs are then used by the classifier on subsequent input to reproduce the expert labels. Descriptors based on histogram statistics, co-occurrence matrices, run length parameters, and fractal measures are typically used to enumerate the features. Artificial neural networks, Bayesian classifiers, support vector machines and *k*-neighborclassifiers are used to classify the features[41–43].

The detection and quantification of pulmonary parenchyma by CALIPER is based on histogram signature mapping techniques trained through expert radiologist consensus assessment of pathologically confirmed datasets obtained through the Lung Tissue Research Consortium (LTRC). The LTRC is a NIH/NHLBI sponsored, multi-site initiative dedicated to helping investigators develop a better understanding of ILD and COPD through development of a repository of clinical data (including demographics, questionnaire responses, medical history), physiologic data, pathological specimens, blood and tissue characterization, CT scan data. The LTRC also provides central expert review of the clinical, radiologic, pathologic data. The data and tissue specimens on the well-characterized subjects in the LTRC are available at no cost to qualified investigators through a standardized process overseen by the NIH/NHLBI (requests can be made at http://www.ltrcpublic.com/data_requests.htm). For our training VOIs, HRCT scans from fourteen subjects were selected from the LTRC repository to create a set of 976 VOIs. The VOIs were selected through independent characterization by four subspecialty thoracic radiologists, with instructions to determine if visual appearance of 70% or more of the given VOI spanning 15×15×15 voxels was normal, contains emphysema or belonged to one of the characteristic ILD parenchymal CT patterns: ground glass opacities (GG), reticular infiltrates (RI), or honeycombing (HC). Based on this criterion 80, 150, 187, 265 and 294 VOIs were determined by consensus agreement selected to represent emphysema, ground glass, honeycombing, normal and reticular infiltrates, respectively.

Quantitative discriminability of a number of pairwise dissimilarity metrics based on the VOI histograms was examined using multi-dimensional scaling (MDS)[44]. Parametric and non-parametric dissimilarity metrics tested to evaluate the optimal dissimilarity included first and second order statistics and measures of effectiveness[45] such as Fechner-Weber contrast measure, target-reference inference ratio, Fisher distance, correlation coefficient, scale invariant normalized mean square error and normalized mutual information. Non-parametric dissimilarity metrics were based on pairwise histogram distances such as Manhattan,

Euclidean, Bhattacharya, Kolmogrov-Smirnoff and Cramer Von Mises Distance (CVM), chi squared distance, Kullback-Liebler divergence, Jeffrey divergence, and histogram intersection[46]. Of all the metrics, MDS representation of CVM (the squared L_2-metric between cumulative density functions) was found to be most consistent with the expert groupings and, consequently, was chosen as the dissimilarity metric used by CALIPER in the automated classification. Figure 1 shows the three dimensional MDS projection for the pairwise CVM dissimilarity measure of the expert-characterized VOI histograms, revealing the natural orderliness with which the MDS representation of CVM metric projects the VOIs to align with expert consensus. Concordance between the consensus radiology labeling (columns) and the affinity propagation based, unsupervised clustering of the pairwise Cramer Von Mises (CVM) dissimilarity metric (rows) for the 976 Volumes of Interest used to train CALIPER (Computer Aided Lung Informatics for Pathology Evaluation and Rating) was assessed using the Kappa statistic. Multi-category analysis such as this requires the evaluation of the degree of agreement category by category and, therefore, a $K \times K$ tables method for agreement as described by Agresti[47] was utilized to assess the agreement index for each category (i.e. the proportion of agreements observed).

Having established the qualitative equivalence of CVM and expert groupings and verified the results with expert visual validation, quantitative equivalence was evaluated using automatic clustering of CVM similarities. This process was utilized to group the VOIs into natural clusters. To create an unbiased stratification of VOIs into natural clusters, affinity propagation[48] was used. Affinity propagation uses message passing to iteratively find clusters, given pair-wise similarities of n-dimensional data. The clustering based on affinity propagation yielded ten natural clusters. In addition to resolving the clusters, this methodology identifies the exemplar that is most ‘central’ to each of the clusters in feature-space. The most ‘central’ within type of each cluster was identified as the fundamental type of its exemplar class. Since more natural clusters were found than classes of visual abnormalities described for ILD, multiple clusters were shown to correspond to some visual classes: 1, 2, 2, 2, and 3 respectively for emphysema, ground glass, honeycombing, normal and reticular.

The local histograms computed from the 15×15×15 neighborhood of each of the parenchymal voxel were compared against the histogram of the 34 exemplars identified in the training phase. CVM dissimilarity measure was used in the comparison and the fundamental type of the exemplar with the least CVM distance was assigned as the parenchymal class of the underlying voxel. The number of voxels belonging to each of the parenchymal classes was calculated across the whole lung and the individual lungs. The voxels identified as vessels were included as normal to account for the total lung volume.

Meaningful representation of the volumetric classification and quantification of abnormalities are key factors in the utility and acceptability of results in clinical practice. Despite the continuing progress in quantitative imaging in many areas of imaging science, the lack of unambiguous visualization with accurate, relevant cues severely hinders the clinical adoption of many potentially useful computational tools. In typical clinical practice, it can be difficult for a radiologist to describe thespecific appearance of the manifestations of ILD on HRCT in textas part of the necessary interpretation and reporting results. The description or graphical representation of the type, extent and anatomical/morphological distribution of abnormalities or computational features of disease in 3D space when there are multiple different classes is even more challenging.

There are examples of advanced summary visualization of complex physiologic or morphologic parameters that have been incorporated into routine medical practice, such as the overlay of multi-color PET or SPECT data on anatomic information through image

fusion and depiction of perfusion data in coronary artery disease[49] and brain maps[50] as iconic glyphs. However, there has been no attempt to visualize multi-variable, spatial, temporal, physiologic and pathological information in lungs as summary glyphs. Towards enabling an unambiguous understanding of the health of a patient's lung as derived from multispecialty data, including pulmonary function tests, demographics and high resolution CT, CALIPER uses a lean visualization approach based on "*thin-slicing*" and semiotic theory of stimuli processing[51]. "Thin slicing" is an information filtering approach that enables rapid decision making by converging to the most critical factors from an overwhelming number of variables. Using this process, information from millions of voxels is succinctly aggregated into a comprehensive and easily comprehensible patho-spatio-temporal glyph that provides an unambiguous representation.

A glyph similar to a radial space filling plot[52] was developed to provide an iconic summary of the volumetric parenchymal classification. The glyph provides a global overview of parenchymal characteristics and lung volumes as well as distribution of the components that facilitates comprehension of the multidimensional source data. As shown in Figure 3, the glyph is partitioned with dark radial lines to illustrate the relative volumes of the left and right lungs and further divided into three regions, each representing the individual lobes or upper/middle/lower lung zones, depending on the type of segmentation performed. The asymmetry, if any, between the left and right lungs can be easily observed in the glyph. The individual lobes span through angles proportional to their respective volumes. Within each lobe, the distribution of the computed parenchymal class is represented by color-coded sectors whose size is proportional to the percentage distribution of that pattern in that lobe. Concentric circles are drawn at 20% intervals to enhance visualization. The area of the glyph is chosen to represent the computed total lung volume. Using predictive equations and subject specific demographics information, the normative lung capacity is calculated using guidelines[53] and is overlaid as a distinctive white ring on the glyph. This provides a succinct comparative overview of the subject vis-a-vis the subject's stratified group. Figure 4 shows a mosaic of the glyphs from 372 CT scans from different subjects from the LTRC cohort. The diseases and parenchymal abnormalities include subjects with a primary diagnosis of ILD including usual interstitial pneumonitis (UIP), nonspecific interstitial pneumonitis (NSIP), and hypersensitivity pneumonitis (HSP), subjects with a primary diagnosis of COPD (various morphological subtypes of emphysema including panlobular emphysema, obstructive airway disease such as bronchiolitis and mixed phenotypes of parenchymal and airway disease), control subjects that do not carry a primary diagnosis of ILD or COPD but may have subclinical or secondary diagnosis of parenchymal disease (such as lung cancer or indeterminate nodules), or subjects with mixed or multiple diagnoses such as combined pulmonary fibrosis and emphysema (CPFE). Accounting for the number of voxels classified, it is worth noting that information from 82 gigabytes of HRCT data has been represented in this mosaic space. Even at this resolution, the glyphs provide a succinct overview of the entire database of LTRC subjects with volumetric HRCT at the time of our analysis, and highlight the ease with which the intra- and inter-subject disease distribution can be made apparent. The regional extent of low attenuation due to air trapping or emphysema corresponds to blue areas in the glyphs and therefore differentiation of morphology such as upper-lobe-predominant emphysema vs. diffuse or panlobular disease can be determined at a glance for each subject. Similarly, the regional distribution of the manifestations of ILD and presence or absence of specific features such as honeycombing can be easily determined on each glyph. Moreover, by quantitatively characterizing the pairwise dissimilarity of the regional distributions of the parenchymal patterns, the glyphs and the underlying CT scan data can be automatically stratified into natural groups towards enabling potential quantitative biomarkers to objectively diagnose pathology, track progression and assess pharmacologic response within and across patients[54,55].

As part of the processing of the parenchymal classifiers for each scan, the CALIPER software also determines the centroid of the dominant cluster for each parenchymal class in each of the lung regions. These positional references are tagged to the respective coded regions of the glyph so as to enable navigation from the glyph to the region within the original dataset computed to demonstrate the characteristic parenchymal class. This linking of the summary results to the source data allows the glyph to be more than a simple representation of quantitative results. Specifically, since the results apparent in the glyph remain linked to regions within original volumetric data, positioning the cursor in the classified region of glyph facilitates the ability to shift between summary overview, more granular interrogation of detail, and visual validation of results with on-demand detailed analysis of the HRCT or classification down to the voxel level. Such functionality provides an optimal explorative paradigm for facile visualization and human understandingof not-so-easily comprehensible multidimensional medical datasets[56,57].

For enhanced visualization, the color code corresponding to the parenchymal classifier is displayed as a transparent overlay on the HRCT data. To facilitate understanding of the quantity of each class within a region, the volume of each class is displayed by hovering over the different colors represented in the glyph. Figures 5 demonstrates how the representation of results in the summary glyph is linked to the underlying three dimensional scan data.

The utility of a summary glyph can be even more powerful when used for evaluation of volumetric data over time. The distribution patterns of the disease and changes in the extent and characteristics of the disease are apparent when the summary glyphs are viewed side by side, and thereby allowing simultaneous interrogation of the quantity of visible pathology, 3D spatial distribution of disease and temporal changes within multiple datasets. For example, Figure 6 is an illustration of glyphs of a subject with progressing interstitial lung disease at four successive time points.

To validate the results of the quantitative analysis, associations between CALIPER measures, PFTs, and radiologists' scores were assessed using Spearman's correlation coefficient. In longitudinal and survival analysis, selected patients evaluated at Mayo Clinic Rochester from January 2000 to December 2010 with IPF as diagnosed according to the latest international consensus guidelines (with or without surgical lung biopsy) and for whom at least two serial HRCT obtained within a 3 to 15 month interval were included. HRCT obtained at the time of acute exacerbation, pneumonia, fluid overload or thromboembolic disease were excluded. Changes in CALIPER measures from between the two CT scans of each subject were calculated for total volume of normal lung, GG, R, HC, total lung volume, total ILD volume (GG+R+HH), and percent ILD. The association of change in CALIPER measures and change in radiologists' assessments of regional semi-quantitative findings and overall impression of progression, regression or stability of disease with survival was assessed using Cox proportional hazards regression. A landmark survival analysis was performed with survival indexed from the date of the second CT. Both univariable and multivariable analyses including gender, pack years, baseline PPFVC, baseline PPDLCO, and time between CTs were performed. Quantitative efficacy of the clustering was established by computing the ANalysis Of SIMilarity (ANOSIM).[58] Similar to the more commonly used ANalysis Of VAriance between groups test (ANOVA), the ANOSIM test is based on comparing distances between groups with distances within groups. The ANOSIM test statistic (R) is a measure of the strength of difference among groups, and in our specific usage this statistic indicates the magnitude difference between clusters. An R value of 1 indicates that the communities completely differ among defined groups; a value of 0 indicates no difference among groups. In all cases, two-sided p-values <0.05 were considered statistically significant.

Results

The natural clusters and the groupings were highly correlated to the consensus groupings of the VOIs determined by experts as shown in the confusion matrix in Table 1. The tightness of the clusters was evaluated by computing the mean dissimilarities between the exemplars and the mean dissimilarities between the candidates of the individual clusters. The mean intra-cluster and inter-exemplar values shown in Figure 2 highlights the compactness of the individual clusters and significant feature space discriminability of the individual exemplars. The combined ANOSIM R value for the five clusters was found to be 0.962 +/− 0.017, with $p < .005$, highlighting the very high level of agreement of candidates within each of the clusters. The results of this exercise attest to the optimality of the chosen dissimilarity metric and clustering technique to automatically and independently group the VOIs into classes derived through expert consensus.

The VOIs within each of the classes were re-clustered independently, again using unsupervised affinity propagation, to obtain the respective class specific fundamental exemplars. This process resulted in a total of 34 fundamental exemplar VOIs corresponding to the visual features of pulmonary parenchyma: 5 for emphysema, 5 for ground glass, 6 for honeycombing, 9 for reticular and 9 for normal parenchyma. These exemplars were subsequently used as the reference VOIs to identify the label that best matches with the 15×15×15 neighborhood centered around each voxel of the test dataset to be classified.

A subset of subjects (N=119) with proven diagnosis of interstitial lung disease was selected from the LTRC database to perform the correlation studies[59,60]. The studies used as training datasets to determine the exemplar/reference ROIs for the classifier were not used in the analysis. Significant correlations were noted between CALIPER characterization and multiple physiologic parameters that are currently accepted as biomarkers for disease severity in ILD. Specifically, percent involvement of reticular infiltrates correlated with significant ($p < 0.001$) changes in 6-minute walk total distance ($r = −0.32$), FVC pre-bronchodilator ($r = −0.63$), DLCO ($r = −0.65$) and TLC ($r = −0.44$). Similarly, significant ($p < 0.001$) inverse correlation existed between lung classified as normal and physiologic tests, including 6-minute walk test ($r = 0.32$), FVC pre-bronchodilator ($r = 0.66$), DLCO ($r = 0.59$) and TLC ($r = 0.56$).

To verify that the automated classification of the abnormalities matched those of expert radiologist description of disease, regional matching to severity and character of disease determined by the interpreting radiologist for the LTRC was performed. Specifically, the LTRC database records semi-quantitative scores of visual abnormalities for each subject with CT data. These scores are coded from 0–4 to respectively represent 0%, 0–25%, 25–50%, 50–75%, and 75–100% of abnormality type in each of 12 regions: the central and peripheral zones of right and left upper, middle and lower lobes (lingula on the left). Spearman's correlation analysis was performed for the individual regions between visual scores and quantitative volumes for Ground Glass (GG), honeycomb (HC), reticular infiltrates (RI), the sum of (GG+HC+RI) and (HC+RI). The regional visual scores correlated significantly ($p < 0.001$) for (GG+HC+RI) ($r = 0.38–0.58$), (HC+RI) ($r = 0.31 – 0.65$) and inversely for normal parenchymal class. Significant correlations were found for GG, HC and RI quantitative scores and corresponding visual scores: GG ($r= 0.19– 0.42$, $p < 0.039$), HC ($r=0.27–0.56$, $p < 0.003$), RI ($r =0.18–0.47$, $p < 0.05$).

Discussion

The potential for quantitative analysis to reliably characterize and quantify parenchymal abnormalities of HRCT in the setting of ILD is enormous. Ideally, computational tools can

yield an objective biomarker thatmay allow for more consistent characterization of disease, with a mapping of specific characteristics and parenchymal abnormalities. Initial investigations on well-characterized data from the LTRC and retrospective analysis of clinical data in subjects with ILD demonstrate correlation with other biomarkers of pulmonary disease such as physiologic testing and also show promise as an independently significant predictor of outcomes. In particular, previously published data with retrospective analysis of a small cohort of 55 subjects with UIP using an earlier version of the CALIPER software showed the potential for quantitative ILD analysis to provide an independent biomarker that can predict survival based. The clinical high-resolution CT scans of these subjects and corresponding pulmonary function test (PFT) data were assessed baseline and follow-up (at least two serial HRCT obtained within a 3 to 15 month interval, with PFT data obtained within 30 days of CT scan). Clinical follow-up of the subjects was reviewed for transplant or mortality, with median transplant-free survival of 2.1 years (range, 1.1 to 3.4) from the second HRCT time point. The HRCT scans were also independently evaluated by subspecialty thoracic radiologists blinded to the clinical data by the same semi-quantitative scale for regional visual abnormalities as the LTRC. The results of this study[61] indicate that interval change in volume of reticular densities (Hazard Ratio (HR) 1.91, P = 0.006), total volume of interstitial abnormalities (HR 1.70, P = 0.003) and percent total interstitial abnormalities (HR 1.52, P = 0.017) as quantified by CALIPER were predictive of survival, even with adjustment for smoking history (pack-years), gender, percent predicted FVC and DLCO and time between HRCT scans. Interestingly, the changes in total extent of ILD and total reticulation scores as assessed semi-quantitatively by radiologists were not found to be predictive of survival, although the overall qualitative assessment of ‘worsening’ by the radiologist was predictive.

Demonstration of progression of disease or response to therapy that may not be obvious to a radiologist would be helpful in clinical management. Similarly, determination of a quantity of disease or characteristic pattern that is independently predictive of mortality could be used to triage patients for transplant or other therapy. Assessment of results for multiple time points in a prospective longitudinal study or examination of previously characterized subjects involved in a longitudinal study will likely be necessary for additional validation of the CALIPER tool. Adequate visualization of quantitative results is also extremely important to real-world utility and acceptance of these results for patient management. A summary that provides unambiguous pathological, spatial and temporal information enables multivariate, multidimensional data to be intelligible. The ability to evaluate the results in context to the original CT data such as demonstrated in the CALIPER glyphs adds to the utility of these results.

The variability of input data to any classification scheme is an inherent problem for any type of quantitative analysis. CT scan parameters that significantly alter the pixel values through edge-enhancing kernels and filters often cause significant differences in the classifier output. This is predictable since the CALIPER process relies on volumetric histogram analysis and uses a training set obtained from non-edge-enhanced HRCT data. Image analytics strategies to understand the nuances of CT reconstruction parameters[62] and consequently, standardization of the techniques used to reconstruct image data across vendors or scanner versions would greatly simplify the task of developing a reliable quantitative analysis tool, and will likely be necessary not only for assessment of ILD but also for any processing of CT that relies on densitometry or histogram analysis.

It is interesting to note that the correlation of CALIPER results with physiologic parameters was generally stronger than with the semi-quantitative radiologist assessment of disease type and severity. Presumably, this analysis is confounded by the lack of ground truth for subjective semi-quantitative findings and the reality that there is inter- and intra- rater

variability between the radiologists preforming interpretations for the LTRC. It is notable that the correlation of the classifier with the radiologist score is similar to the inter-radiologist correlation seen with blinded semi-quantitative scoring performed by multiple expert radiologists[40]. In addition, review of the areas of discrepancy between the automated classifier typically appeared to be in regions that are often confusing to even a human, such as areas of subtle ground glass, course reticular densities or areas of traction bronchiectasis that appear similar to honeycombing or areas of emphysema in subjects with pulmonary fibrosis in which either honeycombing or emphysematous regions could be mis-classified.

With the optimizations in software and employment of modern computational hardware, CALIPER is able to complete segmentation, classification and calculate results for a volumetric HRCT of more than 600 slices in approximately 1 minute. Transitioning these tools from primarily a research role to clinical practice remains a challenge, however. As with all quantitative analysis tools, this transition will require recognition of the legitimacy of results, creation of new workflows and dataflows within clinical systems, and optimization of the technology so that results are available in clinically relevant time. It is hoped that the more robust future validation, acceptance of these results and those of other automated tools being developed throughout the community of medical imaging scientists will soon substantiate the utility of quantitative analysis in clinical practice.

Acknowledgments

This research was partially funded by a contract from the National Institutes of Health HHSN268201100022C, as the Mayo Clinic serves as the Lung Tissue Research Consortium Radiology Center and Dr. Bartholmai is the Principal Investigator. There are no other potential conflicts of interest to disclose.

References

1. Raghu G, Weycker D, Edelsberg J, Bradford WZ, Oster G. Incidence and prevalence of idiopathic pulmonary fibrosis. Am J Respir Crit Care Med. Oct 1; 2006 174(7):810–816. [PubMed: 16809633]
2. Ley B, Collard HR, King TE Jr. Clinical course and prediction of survival in idiopathic pulmonary fibrosis. Am J Respir Crit Care Med. Feb 15; 2011 183(4):431–440. [PubMed: 20935110]
3. Cottin V, Nunes H, Mouthon L, et al. Combined pulmonary fibrosis and emphysema syndrome in connective tissue disease. Arthritis Rheum. Jan; 2011 63(1):295–304. [PubMed: 20936629]
4. Lynch DA, Godwin JD, Safrin S, et al. High-resolution computed tomography in idiopathic pulmonary fibrosis: diagnosis and prognosis. Am J Respir Crit Care Med. Aug 15; 2005 172(4): 488–493. [PubMed: 15894598]
5. Mueller-Mang C, Grosse C, Schmid K, Stiebellehner L, Bankier AA. What every radiologist should know about idiopathic interstitial pneumonias. Radiographics. May-Jun;2007 27(3):595–615. [PubMed: 17495281]
6. Agusti C. American Thoracic Society/European Respiratory Society International Multidisciplinary Consensus Classification of the Idiopathic Interstitial Pneumonias (vol 165, pg 277, 2002). American Journal of Respiratory and Critical Care Medicine. Aug 1; 2002 166(3):426–426.
7. Raghu G, Collard HR, Egan JJ, et al. An Official ATS/ERSARS/ALAT Statement: Idiopathic Pulmonary Fibrosis: Evidence-based Guidelines for Diagnosis and Management. American Journal of Respiratory and Critical Care Medicine. Mar 15; 2011 183(6):788–824. [PubMed: 21471066]
8. Peikert T, Daniels CE, Beebe TJ, Meyer KC, Ryu JH, Phys ACC. Assessment of current practice in the diagnosis and therapy of idiopathic pulmonary fibrosis. Respiratory Medicine. Sep; 2008 102(9):1342–1348. [PubMed: 18621518]
9. Raghu G, Mageto YN, Lockhart D, Schmidt RA, Wood DE, Godwin JD. The accuracy of the clinical diagnosis of new-onset idiopathic pulmonary fibrosis and other interstitial lung disease: A prospective study. Chest. Nov; 1999 116(5):1168–1174. [PubMed: 10559072]
10. Kazerooni EA. High-resolution CT of the lungs. AJR Am J Roentgenol. Sep; 2001 177(3):501–519. [PubMed: 11517038]

11. Misumi S, Lynch DA. Idiopathic pulmonary fibrosis/usual interstitial pneumonia: imaging diagnosis, spectrum of abnormalities, and temporal progression. Proc Am Thorac Soc. Jun; 2006 3(4):307–314. [PubMed: 16738194]
12. Diette GB, Scatarige JC, Haponik EF, Merriman B, Fishman EK. Do high-resolution CT findings of usual interstitial pneumonitis obviate lung biopsy? Views of pulmonologists. Respiration. 2005; 72(2):134–141. [PubMed: 15824522]
13. Lynch DA, Travis WD, Muller NL, et al. Idiopathic interstitial pneumonias: CT features. Radiology. Jul; 2005 236(1):10–21. [PubMed: 15987960]
14. Zerhouni E. High resolution CT for evaluation of diffuse lung disease. CT and MRI of the thorax Churchill Livingstone. 1990:93–105.
15. Costabel, U.; Du Bois, RM.; Egan, JJ. Diffuse parenchymal lung disease. Vol. 36. Karger Publishers; 2007.
16. Kazerooni EA, Martinez FJ, Flint A, et al. Thin-section CT obtained at 10-mm increments versus limited three-level thin-section CT for idiopathic pulmonary fibrosis: Correlation with pathologic scoring. American Journal of Roentgenology. Oct; 1997 169(4):977–983. [PubMed: 9308447]
17. Saketkoo LA, Matteson EL, Brown KK, Seibold JR, Strand V, Dis-Related CT. Developing Disease Activity and Response Criteria in Connective Tissue Disease-related Interstitial Lung Disease. Journal of Rheumatology. Jul; 2011 38(7):1514–1518. [PubMed: 21724725]
18. Kim EY, Lee KS, Chung MP, Kwon OJ, Kim TS, Hwang JH. Nonspecific interstitial pneumonia with Fibrosis: Serial high-resolution CT findings with functional correlation. American Journal of Roentgenology. Oct; 1999 173(4):949–953. [PubMed: 10511155]
19. Wells AU, Desai SR, Rubens MB, et al. Idiopathic pulmonary fibrosis - A composite physiologic index derived from disease extent observed by computed tomography. American Journal of Respiratory and Critical Care Medicine. Apr 1; 2003 167(7):962–969. [PubMed: 12663338]
20. Flaherty KR, Andrei AC, King TE, et al. Idiopathic interstitial pneumonia - Do community and academic physicians agree on diagnosis? American Journal of Respiratory and Critical Care Medicine. May 15; 2007 175(10):1054–1060. [PubMed: 17255566]
21. Watadani T, Sakai F, Johkoh T, et al. Interobserver Variability in the CT Assessment of Honeycombing in the Lungs. Radiology. 2013; 266(3):936–944. [PubMed: 23220902]
22. Garland LH. On the scientific evaluation of diagnostic procedures. Radiology. Mar; 1949 52(3): 309–328. [PubMed: 18113241]
23. Tuddenham WJ. Roentgen image perception–a personal survey of the problem. Radiol Clin North Am. Dec; 1969 7(3):499–501. [PubMed: 5382261]
24. Kundel HL. History of research in medical image perception. J Am Coll Radiol. Jun; 2006 3(6): 402–408. [PubMed: 17412094]
25. Flaherty KR, King TE, Raghu G, et al. Idiopathic interstitial pneumonia - What is the effect of a multidisciplinary approach to diagnosis? American Journal of Respiratory and Critical Care Medicine. Oct 15; 2004 170(8):904–910. [PubMed: 15256390]
26. Sverzellati N, Devaraj A, Desai SR, Quigley M, Wells AU, Hansell DM. Method for Minimizing Observer Variation for the Quantitation of High-Resolution Computed Tomographic Signs of Lung Disease. Journal of Computer Assisted Tomography. Sep-Oct;2011 35(5):596–601. [PubMed: 21926855]
27. Lynch DA. Quantitative CT of fibrotic interstitial lung disease. Chest. Mar; 2007 131(3):643–644. [PubMed: 17356073]
28. Delorme S, KellerReichenbecher MA, Zuna I, Schlegel W, vanKaick G. Usual interstitial pneumonia - Quantitative assessment of high-resolution computed tomography findings by computer-assisted texture-based image analysis. Investigative Radiology. Sep; 1997 32(9):566–574. [PubMed: 9291045]
29. Galban CJ, Han MLK, Boes JL, et al. Computed tomography-based biomarker provides unique signature for diagnosis of COPD phenotypes and disease progression. Nature Medicine. Nov; 2012 18(11):1711–+.
30. Depeursinge A, Iavindrasana J, Hidki A, et al. Comparative Performance Analysis of State-of-the-Art Classification Algorithms Applied to Lung Tissue Categorization. Journal of Digital Imaging. Feb; 2010 23(1):18–30. [PubMed: 18982390]

31. Uppaluri R, Hoffman EA, Sonka M, Hunninghake GW, McLennan G. Interstitial lung disease - A quantitative study using the adaptive multiple feature method. American Journal of Respiratory and Critical Care Medicine. Feb; 1999 159(2):519–525. [PubMed: 9927367]
32. Shyu CR, Brodley CE, Kak AC, Kosaka A, Aisen AM, Broderick LS. ASSERT: A physician-in-the-loop content-based retrieval system for HRCT image databases. Computer Vision and Image Understanding. Jul-Aug;1999 75(1–2):111–132.
33. Hu SY, Hoffman EA, Reinhardt JM. Automatic lung segmentation for accurate quantitation of volumetric X-ray CT images. Ieee Transactions on Medical Imaging. Jun; 2001 20(6):490–498. [PubMed: 11437109]
34. Sato Y, Westin CF, Bhalerao A, et al. Tissue classification based on 3D local intensity structures for volume rendering. Ieee Transactions on Visualization and Computer Graphics. Apr-Jun;2000 6(2):160–180.
35. Uppaluri R, Hoffman EA, Sonka M, Hartley PG, Hunninghake GW, McLennan G. Computer recognition of regional lung disease patterns. American Journal of Respiratory and Critical Care Medicine. Aug; 1999 160(2):648–654. [PubMed: 10430742]
36. Uchiyama Y, Katsuragawa S, Abe H, et al. Quantitative computerized analysis of diffuse lung disease in high-resolution computed tomography. Medical Physics. Sep; 2003 30(9):2440–2454. [PubMed: 14528966]
37. Hoffman EA, Reinhardt JM, Sonka M, et al. Characterization of the interstitial lung diseases via density-based and texture-based analysis of computed tomography images of lung structure and function. Academic Radiology. Oct; 2003 10(10):1104–1118. [PubMed: 14587629]
38. Chabat F, Yang GZ, Hansell DM. Obstructive lung diseases: Texture classification for differentiation at CT. Radiology. Sep; 2003 228(3):871–877. [PubMed: 12869685]
39. Kim KG, Goo JM, Kim JH, et al. Computer-aided diagnosis of localized ground-glass opacity in the lung at CT: Initial experience. Radiology. Nov; 2005 237(2):657–661. [PubMed: 16192320]
40. Zavaletta VA, Bartholmai BJ, Robb RA. High resolution multidetector CT-aided tissue analysis and quantification of lung fibrosis. Academic Radiology. Jul; 2007 14(7):772–787. [PubMed: 17574128]
41. van Ginneken B, Katsuragawa S, Romeny BMT, Doi K, Viergever MA. Automatic detection of abnormalities in chest radiographs using local texture analysis. Ieee Transactions on Medical Imaging. Feb; 2002 21(2):139–149. [PubMed: 11929101]
42. Xu Y, van Beek EJR, Yu HJ, Guo JF, McLennan G, Hoffman EA. Computer-aided classification of interstitial lung diseases via MDCT: 3D adaptive multiple feature method (3D AMFM). Academic Radiology. Aug; 2006 13(8):969–978. [PubMed: 16843849]
43. Arzhaevaa Y, Prokop M, Tax DMJ, De Jong PA, Schaefer-Prokop CM, van Ginneken B. Computer-aided detection of interstitial abnormalities in chest radiographs using a reference standard based on computed tomography. Medical Physics. Dec; 2007 34(12):4798–4809. [PubMed: 18196808]
44. Trevor, Cox; Cox, M. Multidimensional Scaling. Chapman & Hall/CRC; 2001.
45. Sadjadi, F. Paper presented at: Geoscience and Remote Sensing Symposium, 2003 IGARSS'03. Proceedings. 2003 IEEE International. 2003. Measures of effectiveness and their use in comparative image fusion analysis.
46. Elena, Deza; Deza, M. Dictionary of Distances. Elsevier; 2006.
47. Agresti, A. Categorical data analysis. 3. Hoboken, NJ: Wiley; 2013.
48. Frey BJ, Dueck D. Clustering by passing messages between data points. Science. Feb 16; 2007 315(5814):972–976. [PubMed: 17218491]
49. Oeltze, S.; Kuß, A.; Grothues, F.; Hennemuth, A.; Preim, B. Paper presented at: Proceedings of the Eighth Joint Eurographics/IEEE VGTC conference on Visualization. 2006. Integrated visualization of morphologic and perfusion data for the analysis of coronary artery disease.
50. Domin, M.; Langner, S.; Hosten, N.; Linsen, L. Visualization in medicine and life sciences. Springer; 2008. Direct glyph-based visualization of diffusion MR data using deformed spheres; p. 185-204.
51. Ward MO. A taxonomy of glyph placement strategies for multidimensional data visualization. Information Visualization. 2002; 1(3–4):194–210.

52. Andrews, K.; Heidegger, H. Paper presented at: Proc of IEEE Infovis' 98 late breaking Hot Topics. 1998. Information slices: Visualising and exploring large hierarchies using cascading, semi-circular discs.

53. Crapo RO, Morris AH, Gardner RM. Reference Spirometric Values Using Techniques and Equipment That Meet Ats Recommendations. American Review of Respiratory Disease. 1981; 123(6):659–664. [PubMed: 7271065]

54. Raghunath, S.; Rajagopalan, S.; Karwoski, R.; Bartholmai, B.; Robb, R. Paper presented at: Biomedical Imaging (ISBI), 2012 9th IEEE International Symposium. 2012. Quantitative image analytics for stratified pulmonary medicine.

55. Raghunath, S.; Rajagopalan, S.; Karwoski, RA.; Bartholmai, BJ.; Robb, RA. Medical Image Computing and Computer-Assisted Intervention–MICCAI 2011. Springer; 2011. Referenceless stratification of parenchymal lung abnormalities; p. 223-230.

56. Ropinski T, Preim B. Taxonomy and usage guidelines for glyph-based medical visualization. Proc Simulation and Visualization (SimVis). 2008:121–138.

57. Shneiderman, B. Paper presented at: Visual Languages, 1996. Proceedings., IEEE Symposium. 1996. The eyes have it: A task by data type taxonomy for information visualizations.

58. Anderson MJ. A new method for non-parametric multivariate analysis of variance. Austral Ecol. Feb; 2001 26(1):32–46.

59. Raghunath S, Moua T, Segovis C, et al. Correlation of Automated Quantitative Measures of Interstitial Lung Disease (ILD) Using CALIPER With Semiquantitative Visual Radiology Scores. CHEST Journal. 2011; 140(4_MeetingAbstracts):1042A–1042A.

60. Raghunath S, Moua T, Segovis C, et al. Correlation of Quantitative Lung Tissue Characterization as Assessed by CALIPER With Pulmonary Function and 6-Minute Walk Test. CHEST Journal. 2011; 140(4_MeetingAbstracts):1040A–1040A.

61. Maldonado F, Moua T, Rajagopalan S, et al. Automated quantification of radiologic patterns predicts survival in idiopathic pulmonary fibrosis. European Respiratory Journal. 2013

62. Raghunath, S.; Rajagopalan, S.; Karwoski, R., et al. Paper presented at: Biomedical Imaging (ISBI), 2013 10th IEEE International Symposium. 2013. Landscaping the Effect of CT Reconstruction Parameters: Robust Interstitial Pulmonary Fibrosis Quantitation.

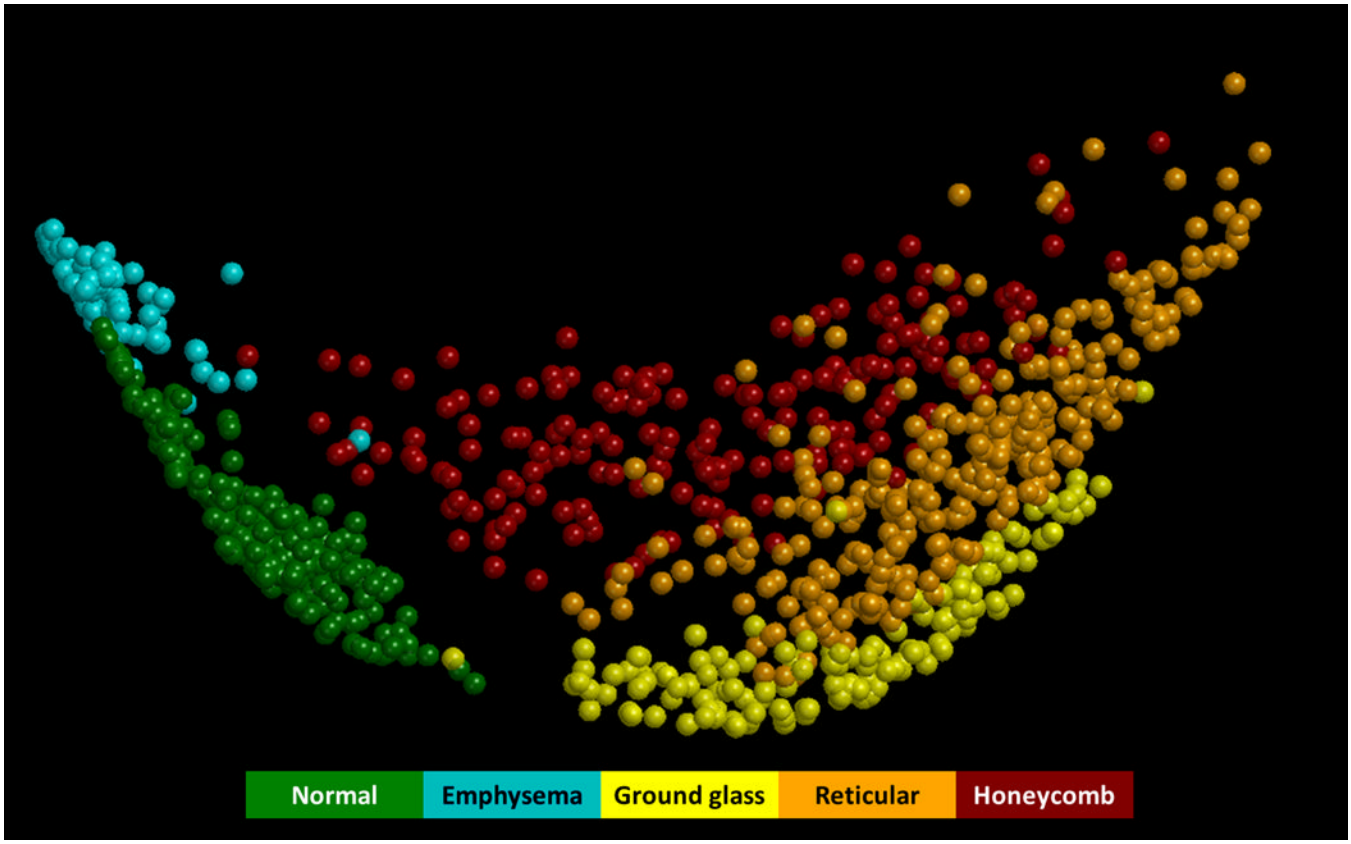

Figure 1.
2D projection of the 3D MDS for the pairwise CVM dissimilarity of the 976 color-coded VOIs used to train CALIPER.

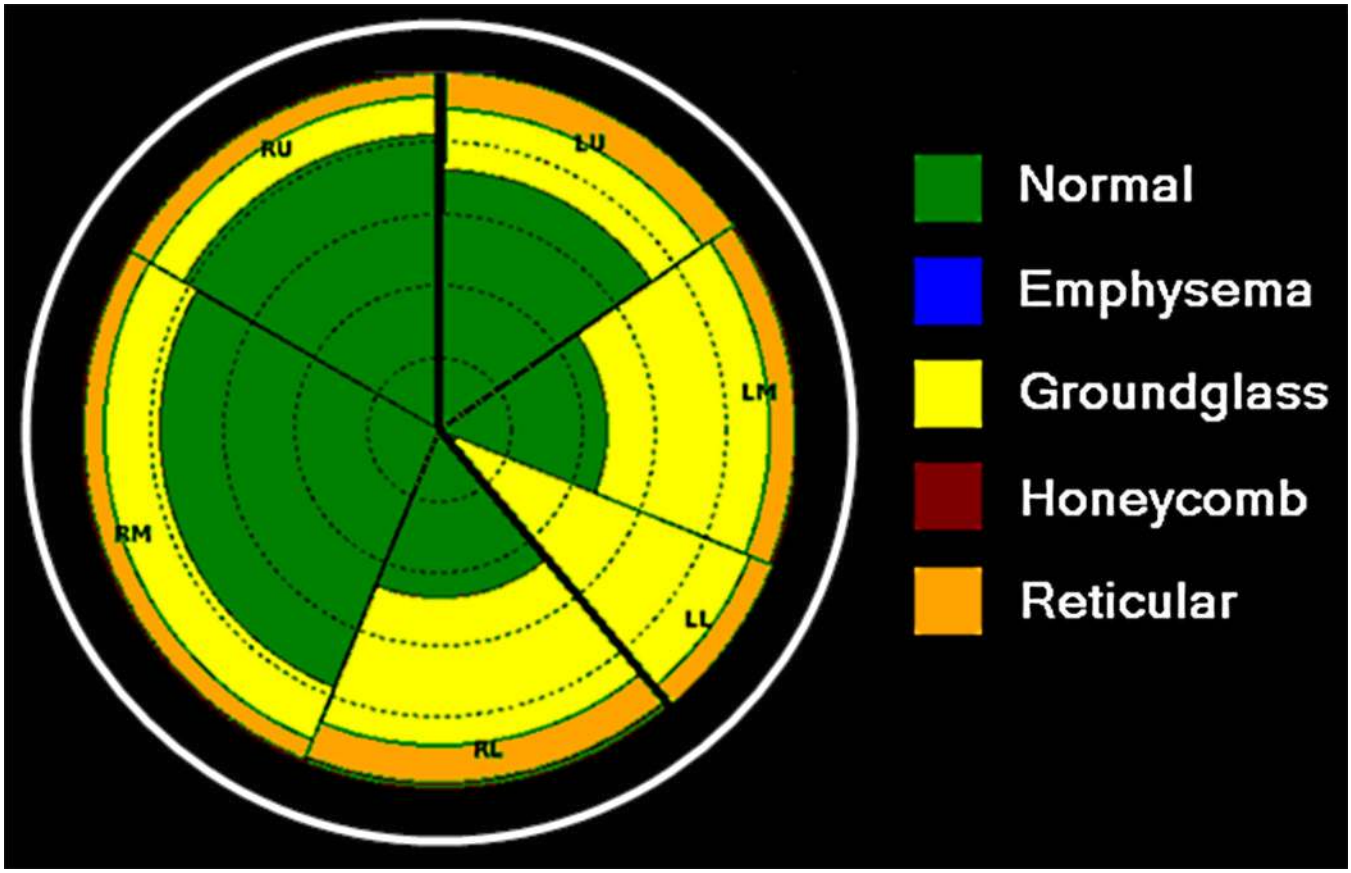

Figure 2.
Representative glyph provides summary of distribution for 5 characteristic CT patterns (color-coded) in lung parenchyma. The first letter (R/L) denotes the right and left lung, the second letter (U/M/L) denotes, respectively, the upper, middle, and lower lung zones defined on the basis of the automatic detection of carina. The radius of the glyph is proportional to the lung volume. The white outer circle denotes the predicted total lung capacity of the subject.

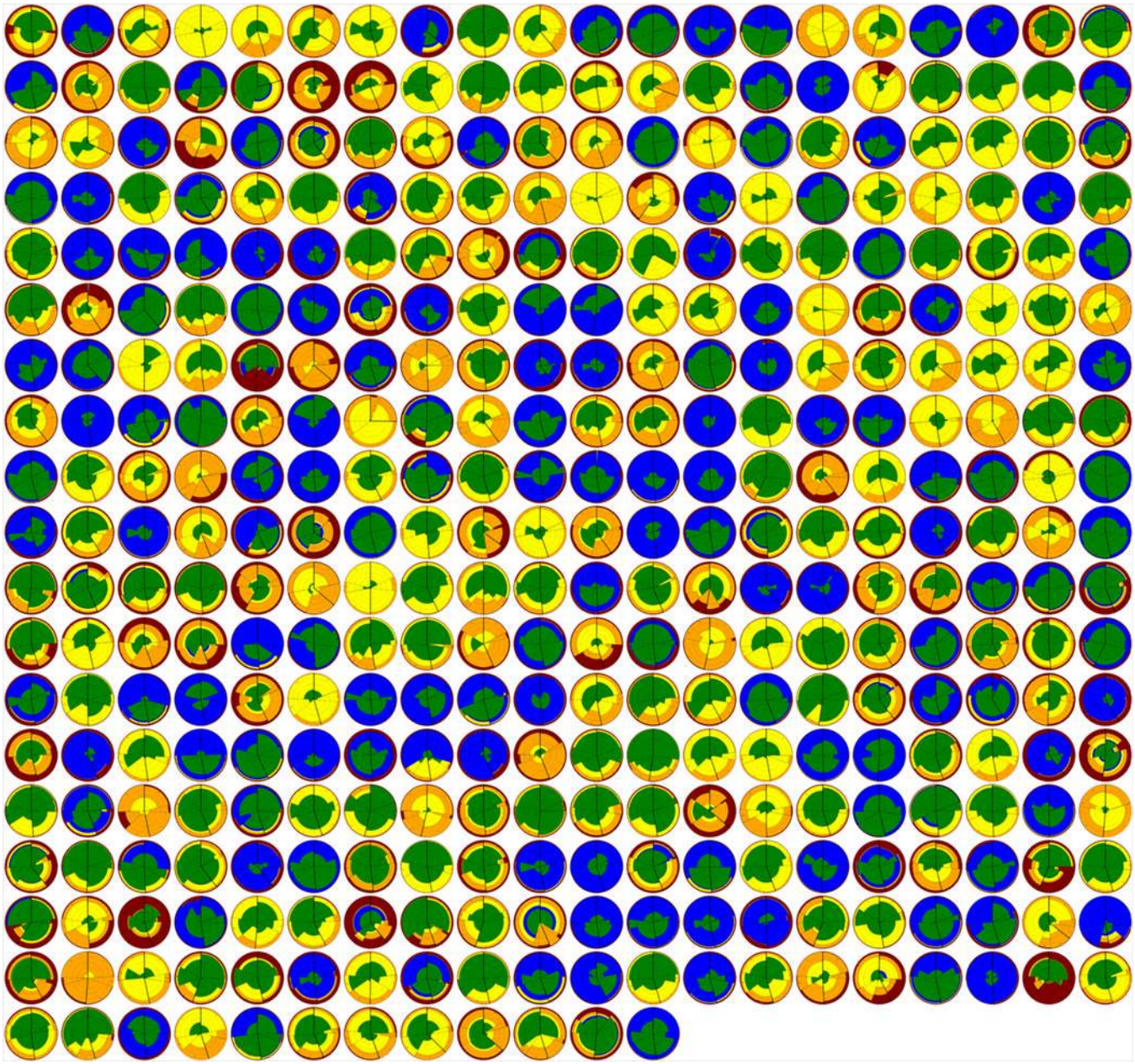

Figure 3.
The montage of glyphs for CT scans of 372 LTRC subjects demonstrates the spectrum of parenchymal abnormalities. Subjects in the LTRC shown here represent a variety of ILD, COPD, and mixed parenchymal and/or airway diseases including UIP, nonspecific interstitial pneumonitis, hypersensitivity pneumonitis, emphysema, bronchiolitis, and combined pulmonary fibrosis and emphysema. COPD subjects generally have more blue (representing low-attenuation/emphysema regions), whereas ILD or mixed parenchymal diseases demonstrate more yellow, orange, and red (GG, RI, and HC, respectively). There are few, if any, truly "normal" subjects in the LTRC database, and therefore only a minority of glyphs exhibit predominantly green throughout all regions.

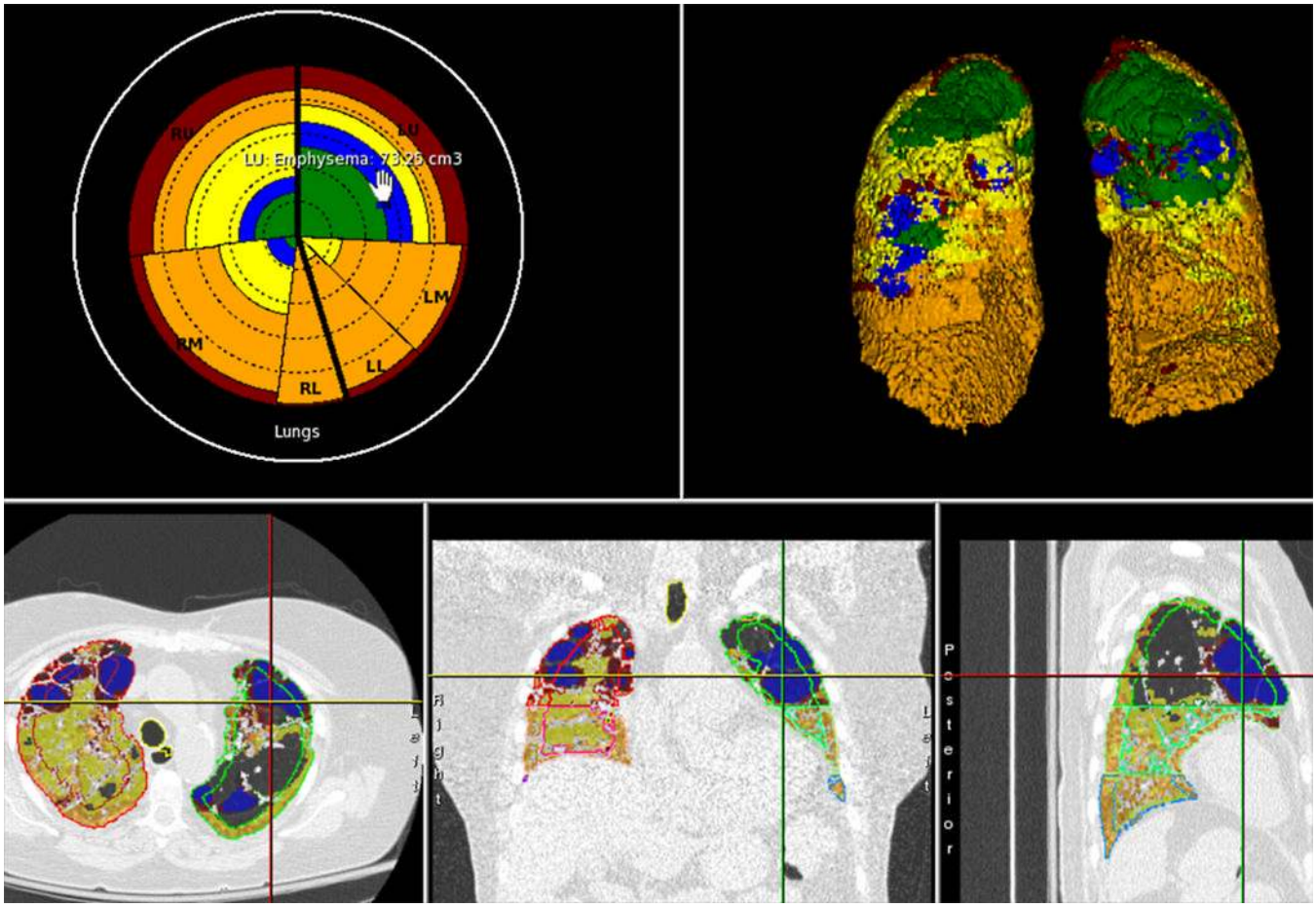

Figure 4.

Location tagging of the summary glyph (top) to the underlying 3D scan data (bottom). In this example, clicking on the emphysema (blue)—left upper (LU)—region of the glyph reveals the triplanar view of the most characteristic emphysema region in the left upper lobe. The first letter (R/L) indicates the right and left lung, the second letter (U/M/L) denotes, respectively, the upper, middle, and lower lung zones.

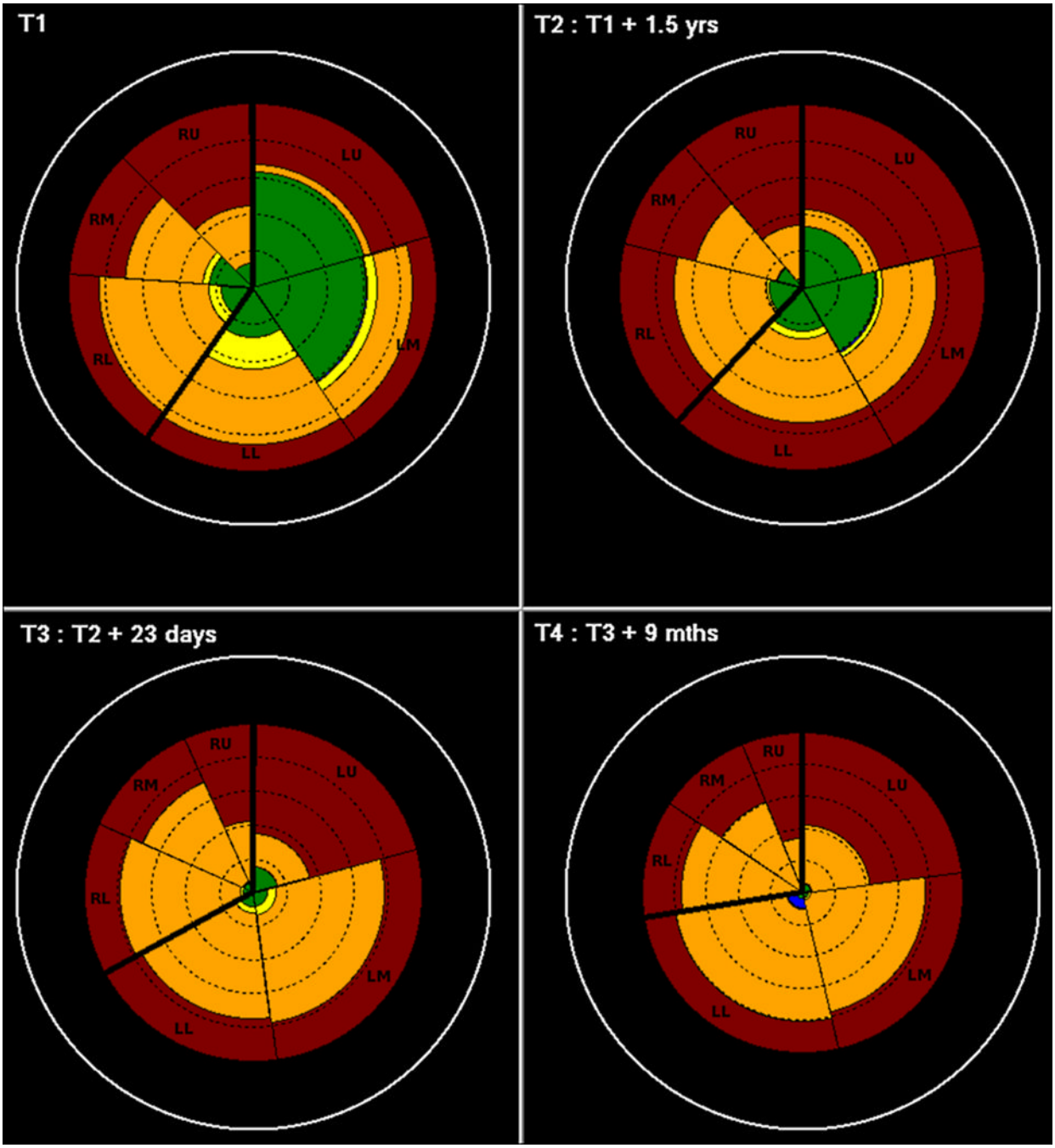

Figure 5.
Patho-spatio-temporal glyph illustration for a subject with progressing fibrotic ILD. The decreasing lung volume (radius of the glyph) with respect to the outer white circle (predicted total lung capacity), loss of normal parenchyma, and increasing proportion of HC and RI densities in the volume is apparent. Volume loss in the right lung is relatively greater than in the left. The first letter (R/L) indicates the right and left lung, the second letter (U/M/L) denotes, respectively, the upper, middle, and lower lung zones.

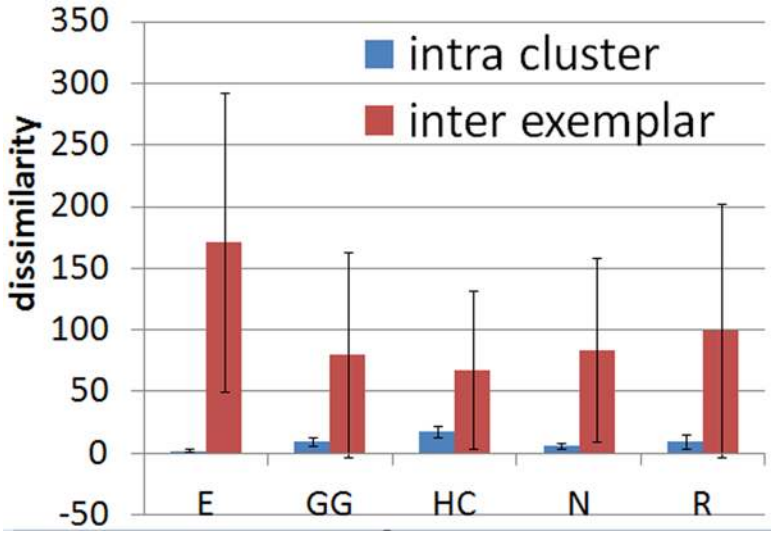

Figure 6.

Graph revealing that the mean pairwise CVM dissimilarity distances between the pattern-specific exemplars are significantly higher than the mean dissimilarity distances within the VOIs belonging to the individual cluster. The high interexemplar and low intracluster values signify high cluster validity. Quantitative efficacy of the clustering was established by computing the similarity statistic (R) by the ANOSIM method. The combined R for the 5 clusters was found to be 0.962±0.017, with $P<0.005$. E indicates emphysema; N, normal.

Table 1

Concordance Between the Consensus Radiology Labeling (Columns) and the Affinity Propagation-based, Unsupervised Clustering of the Pairwise CVM Dissimilarity Metric (Rows) for the 976 VOIs Used to Train CALIPER

	E	GG	HC	N	RI
Emphysema (E)	77	0	0	3	0
GG	0	137	1	2	10
HC	1	11	148	7	20
Normal (N)	0	0	0	265	0
RI	0	16	32	0	246

The strength of agreement as evaluated using κ statistics was 0.863 (95% confidence interval, 0.838–0.888).