Online supplement

1. Texture Feature extraction
1.1 Training set generation
High resolution CT scans from fourteen patients with moderate diffuse pulmonary lung disease was selected from the Lung Tissue Research Consortium repository to create a training set of 976 volumes of Interest (VOI). The VOIs were selected through consensus of four expert radiologists such that more than 70% of the region spanning 15x15x15 voxels of VOIs belong to one of the primal lung parenchymal CT patterns- normal, emphysema, ground glass, reticular and honey comb. Based on this criterion 80, 150, 187, 265 and 294 VOIs were selected to represent emphysema, ground glass, honey comb, normal and reticular forms respectively.

1.2 Visualization of VOI similarities
Quantitative discriminability of a number of pairwise similarity metrics based on the VOI histograms was examined using Multi-dimensional scaling (MDS) [1]. Parametric and non-parametric similarity metrics supported in Analyze software version 10.0 (Mayo Clinic, Rochester) were used in this exercise. Parametric metrics included first and second order statistics and measures of effectiveness 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 similarity metrics were based on histogram distances such as Manhattan, Euclidean, Bhattacharya, Kolmogrov-Smirnoff and Cramer
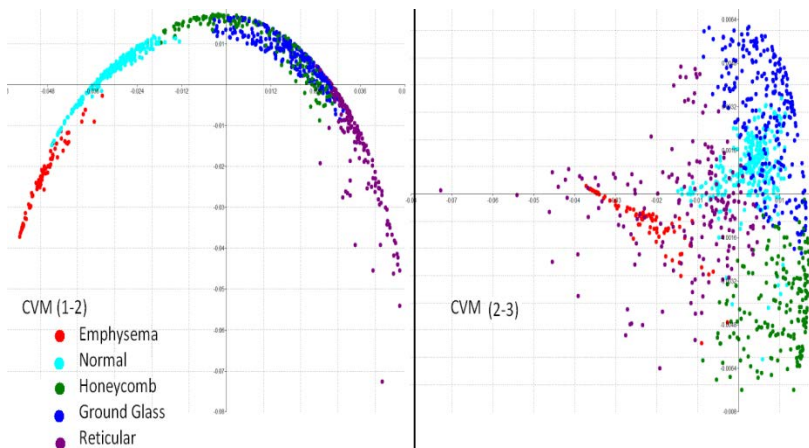
Figure 1. VOI distribution in MDS space for CVM metric. The distribution reveals the correlation of CVM with expert consensus.

Von Mises Distance (CVM), chi squared distance, Kullback-Liebler divergence, Jeffrey divergence, and histogram intersection [2]. 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 similarity metric in the automated classification. Figure 1 shows the axis1-axis2 (1-2) and 2-3 MDS projections for CVM, revealing the natural orderliness with which the metric projects the VOIs to align with expert consensus. The honeycomb and groundglass features overlapping in the 1-2 projection are sufficiently separated in the 2-3 projection.

1.3 Assessing separability quality of CVM similarity metric
Having visually established the statistical equivalence of CVM and expert groupings, quantitative equivalence was evaluated using automatic clustering of CVM similarities and hence the underlying VOIs into natural clusters. To create an unbiased stratification of VOIs into natural clusters, affinity propagation [3] was used. Affinity propagation uses message passing to iteratively find clusters given pair-wise similarities of n-dimensional data. In addition to resolving the clusters, it identifies the exemplar that is most 'central' to each of the clusters.

Clustering based on affinity propagation yielded ten natural clusters. The primal type of each cluster was identified as the primal type of its exemplar. The number of clusters computed was 1, 2, 2, 2, and 3 respectively for emphysema, ground glass, honey combing, normal and reticular. The natural clusters and the groupings were highly correlated to the consensus groupings of experts as shown in the confusion matrix in Table 1. Figure 2 shows mean intra cluster and inter exemplar CVM values for this clustering.

Table 1. Five class confusion matrix; computed class (rows) versus actual class (columns)

	E	GG	HC	N	R
Emphysema (E)	77	0	0	3	0
GroundGlass(GG)	0	137	1	2	10
Honeycombing (HC)	1	11	148	7	20
Normal (N)	0	0	0	265	0
Reticular (R)	0	16	32	0	246
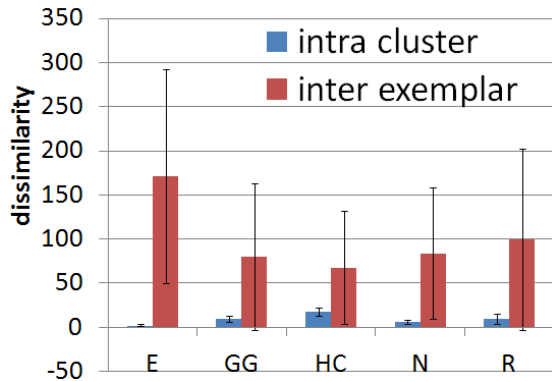
Figure 2. Mean intra cluster and inter exemplar CVM values for the computed class.

1.4 Identification of exemplar VOIs
The VOIs belonging to each of the primal forms were clustered independently using affinity propagation to obtain the respective class specific exemplars. A total of 34 exemplars spanning across the landscapes of emphysema (exemplars E_emph= 5), ground glass (E_groundglass = 5), honey combing (E_honeycombing = 6), normal (E_normal = 9), and reticular (E_reticular = 9). These exempars were collectively used subsequently as reference VOIs to identify the label that best matches with the 15x15x15 neighborhood centered around each of the voxels of the to be classified dataset.

2. Data Processing
2.1 Lung Segmentation
The segmentation of the lungs was achieved using an adaptive density-based morphology approach involving optimal thresholding to identify low density fields in the scans, region growing and void filling [4]. The optimal threshold includes the lung, the air outside the body and other air cavities within the body. The lungs were isolated through connected component analysis. Three dimensional connected components that touch the edge of the volume is assumed

to be air and eliminated. Small connected components that make up for less than one percent of the scan data are also discarded. Three dimensional hole filling was used to fill the lung cavities created by the elimination of normal blood vessels during the thresholding process.

2.2 Airway extraction

The airways of the bronchial tree were automatically segmented by iterative application of increasingly restrictive constraints to a thresholding and 3D region growing process that results in the most complete airway extraction while excluding other low density lung regions such as emphysematous regions, honeycombing or other gas-filled structures such as the stomach or colon. Airway extraction was first attempted utilizing the 1x1x3 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 exceeds 50, the segmentation is assumed to include regions extraneous to airways. In such cases, the segmentation was repeated using more restrictive parameters such as 6 neighborhood connectvity and -960 HU thresholding.

2.3 Lung Separation

In the event the left and right lung continues to be connected after the above steps, morphological operations are applied to separate the lungs. Using morphological erosion iteratively, layers of edge voxels are removed from the segmented lungs so as to break the connectivity of the lungs. After separating the lungs, the removed edge voxels are reaassigned using conditional dilation. This step assigns the removed edge voxels to the closest lung.

The lung borders are subjected to a final smoothing step using morphological closing with a 23x23x5 elliptical structuring element so as to close the fissures along the mediastinum created by pulmonary arteries and veins.

2.4 Vessel Extraction

Pulmonary vessels were extracted using an optimized multi-scale tubular structure enhancement filter based on the eigenvalues of the Hessian matrix [5]. These filters calculate 2nd-order derivatives within the neighborhood of each of the lung voxel. The eigen values of the Hessian matrix constructed from the derivatives are analyzed to determine the likelihood of the underlying voxel belonging to a tubular structure and hence a vessel.

2.5 Parenchymal Classification

The local histograms computed from the 15x15x15 neighborhood of each of the parenchymal voxel were compared against the histogram of the 34 exemplars identified in the training phase. CVM similarity measure was used in the comparison and the primal type of the exemplar with the least CVM distance was assigned as the tissue type of the underlying voxel. The number of voxels belonging to each of the tissue types was calculated across the whole lung and the individual lungs. The voxels identifed as vessels were included as normal to account for the total lung volume.

3. Variations in Parenchymal Classification due to slice thickness

The scans in this study were acquired using GE (Model: LightSpeed Ultra; BONE kernel) and Siemens (Model Sensation 64; B46 kernel) scanners. To assess the efficacy of slice thickness on the parenchymal classification, one representative scan from each of the scanners (slice thickness 0.75 mm) was reconstructed at slice thicknesses of 1.0, 2.0 and 5.0 mm. Total lung volume and the percentage distribution of the classifications across the different tissue types were computed for each of the reconstructions along with their mean and standard deviation. The results shown in Table 2 highlight the robustness of the proposed segmentation and classification algorithm.

Table 2. Effect of slice thickness on Parenchymal Classification

Scanner		Lung Volume (cm^3)	Percentage Distribution of Tissue type				
			E	GG	HC	N	R
Siemens (B46)	Mean	2454	1.52	42.46	6.87	42.99	6.16
	SD	36.86	0.21	3.16	1.55	1.44	0.81
	Max	2510	1.86	48.11	8.32	44.54	6.86
	Min	2409	0.94	40.43	4.81	41.14	4.98
GE (BONE)	Mean	3254	1.16	21.13	6.17	58.07	13.46
	SD	43.63	0.06	0.21	1.78	2.89	1.35
	Max	3308	1.20	21.32	8.9	60.01	15.81
	Min	3216	1.05	20.85	4.8	53.24	12.59

	Interrater and CALIPER	Emphysema (C co-efficient)	GG (C co-efficient)	Reticular (C co-efficient)	Honeycombing (C co-efficient)
LUR	R1 v R2	0.85252	0.63797	0.71755	0.50428
	R1 v C	0.23461	0.38949	0.4687	0.35905
	R2 v C	0.38287	0.3646	0.46901	0.12871
LUC	R1 v R2	0.85104	0.63784	0.55488	0.59963
	R1 v C	0.32491	0.62873	0.46782	0.18836
	R2 v C	0.45728	0.44528	0.30168	0.04509
LMR	R1 v R2	0.4808	0.48179	0.63481	0.37842
	R1 v C	0.05654	0.35226	0.37504	0.29789
	R2 v C	0.0869	0.25056	0.4531	0.12166
LMC	R1 v R2	0.68719	0.41667	0.38577	0.71362
	R1 v C	0.0425	0.55897	0.33621	0.28671
	R2 v C	0.20501	0.33021	0.11283	0.20993
LLR	R1 v R2	-0.03267	0.63481	0.60573	0.63356
	R1 v C	0.1564	0.52663	0.1786	0.22455
	R2 v C	0.24615	0.39741	0.51825	0.24171
LLC	R1 v R2	0.25196	0.64406	0.4737	0.4044
	R1 v C	0.04104	0.46969	0.20685	0.11934
	R2 v C	0.37878	0.46785	0.38032	0.25313
RUR	R1 v R2	0.7443	0.72579	0.72088	0.43391
	R1 v C	0.57423	0.45085	0.57888	0.38095
	R2 v C	0.62653	0.3911	0.52441	0.2865
RUC	R1 v R2	0.7375	0.65525	0.46138	0.32624
	R1 v C	0.35035	0.52797	0.43895	0.03146
	R2 v C	0.47099	0.50801	0.25917	0.10404
RMR	R1 v R2	0.48591	0.6413	0.5329	0.51458
	R1 v C	0.05293	0.39344	0.5236	0.14914
	R2 v C	0.08623	0.22883	0.56611	-0.02266
RMC	R1 v R2	0.58814	0.62044	0.40212	0.63832
	R1 v C	0.17734	0.44335	0.33158	-0.10696
	R2 v C	0.1082	0.24381	0.48304	-0.14701
RLR	R1 v R2	-	0.7133	0.53547	0.60065
	R1 v C	-	0.51849	0.13567	0.32774
	R2 v C	0.28264	0.39933	0.39984	0.38725
RLC	R1 v R2	0.56656	0.74208	0.4985	0.58286
	R1 v C	0.16624	0.55298	0.20317	0.09838
	R2 v C	0.16773	0.44171	0.3283	0.27642

Table 3) Interrater and CALIPER Coefficients at Time Point I
Bold= Significant P value < 0.05
R1= radiologist 1
R2= radiologist 2
C= CALIPER

LUR=left upper rind; LUC=left upper core; LMR=left middle rind; LMC=left middle core; LLR=left lower rind; LLC=left lower core
RUR=right upper rind; RUC=right upper core; RMR=right middle rind; RMC=right middle core; RLR=right lower rind; RLC=right lower core

	Interrater and CALIPER	Emphysema (C co-efficient)	GG (C co-efficient)	Reticular (C co-efficient)	Honeycombing (C co-efficient)
LUR	R1 v R2	0.91606	0.64173	0.76627	0.56596
	R1 v C	0.37631	0.42806	0.57533	0.35414
	R2 v C	0.4574	0.47819	0.55001	0.16902
LUC	R1 v R2	0.93615	0.60319	0.59488	0.59963
	R1 v C	0.2925	0.57454	0.52668	0.22579
	R2 v C	0.37126	0.54881	0.37741	0.21596
LMR	R1 v R2	0.4808	0.51026	0.6663	0.39746
	R1 v C	0.02271	0.40685	0.42528	0.28657
	R2 v C	0.29247	0.40319	0.48416	0.25749
LMC	R1 v R2	0.68719	0.45498	0.46152	0.71362
	R1 v C	0.16676	0.51016	0.39944	0.05733
	R2 v C	0.29361	0.38401	0.2016	0.18215
LLR	R1 v R2	-0.03267	0.67961	0.69764	0.6375
	R1 v C	0.19485	0.55807	0.24123	0.28339
	R2 v C	0.26641	0.48119	0.43839	0.30118
LLC	R1 v R2	0.32588	0.70015	0.5123	0.40445
	R1 v C	0.20658	0.50803	0.24055	0.25988
	R2 v C	0.15144	0.54827	0.28236	0.2883
RUR	R1 v R2	0.80066	0.7326	0.70122	0.43393
	R1 v C	0.37339	0.43539	0.56617	0.36016
	R2 v C	0.45618	0.4994	0.52987	0.30365
RUC	R1 v R2	0.75302	0.63551	0.49167	0.32624
	R1 v C	0.14634	0.56961	0.44906	0.08566
	R2 v C	0.33717	0.63928	0.33926	0.28116
RMR	R1 v R2	0.48591	0.66096	0.64307	0.50356
	R1 v C	0.01801	0.58193	0.5454	0.3022
	R2 v C	0.04633	0.47826	0.51455	0.14477
RMC	R1 v R2	0.51422	0.65828	0.52586	0.72038
	R1 v C	0.20847	0.63905	0.33943	-0.03662
	R2 v C	0.19771	0.55185	0.44174	-0.08898
RLR	R1 v R2	-	0.71727	0.63742	0.58207
	R1 v C	-	0.58931	0.30874	0.35089
	R2 v C	0.31523	0.52403	0.3342	0.39843
RLC	R1 v R2	0.48591	0.77129	0.57176	0.58286
	R1 v C	0.21037	0.57435	0.23822	0.2463
	R2 v C	0.22862	0.5939	0.23269	0.33479

Table 4) Interrater and CALIPER Coefficients at Time Point 2

Bold= Significant P value < 0.05
R1= radiologist 1
R2= radiologist 2

C= CALIPER

LUR=left upper rind; LUC=left upper core; LMR=left middle rind; LMC=left middle core; LLR=left lower rind; LLC=left lower core
RUR=right upper rind; RUC=right upper core; RMR=right middle rind; RMC=right middle core; RLR=right lower rind; RLC=right lower core

Table 5. Survival analysis of CALIPER measurement using Time Point 2

Difference	HR	95% CI		P
Total emphysema†	1.46	0.83	2.57	0.18
Percent emphysema†	1.31	0.84	2.04	0.23
Total ground glass†	1.23	0.86	1.76	0.27
Percent ground glass†	1.17	0.83	1.64	0.37
Total reticular†	3.0	1.65	5.45	<0.001
Percent reticular†	2.27	1.39	3.7	0.001
Total honeycombing†	0.95	0.71	1.26	0.70
Percent honeycombing†	0.95	0.72	1.25	0.71
Total ILD†	1.54	1.06	2.25	0.02
Percent ILD†	1.42	0.98	2.06	0.066

† HR's are for a standard deviation change

Analysis adjusts for gender, pack years, percent predicted FVC at time2, percent predicted DLCO at time 2, and time between HRCTs.

Ref=reference

Table 6. Adjusted Survival analysis of CALIPER measurements using time point 2

Difference	HR	95% CI		P
Total emphysema†	1.46	0.83	2.57	0.18
Percent emphysema†	1.31	0.84	2.04	0.23
Total ground glass†	1.23	0.86	1.76	0.27
Percent ground glass†	1.17	0.83	1.64	0.37
Total reticular†	3.0	1.65	5.45	<0.001
Percent reticular†	2.27	1.39	3.7	0.001
Total honeycombing†	0.95	0.71	1.26	0.70
Percent honeycombing†	0.95	0.72	1.25	0.71
Total ILD†	1.54	1.06	2.25	0.02
Percent ILD†	1.42	0.98	2.06	0.066
Emphysema radiologist 1†	1.92	0.05	74.97	0.73
Emphysema radiologist 2†	0.1	0.01	0.78	0.028
GG radiologist 1†	1.41	0.98	2.03	0.066
GG radiologist 2†	1.30	0.97	1.74	0.086
Honeycombing radiologist 1†	1.0	0.68	1.47	0.99
Honeycombing radiologist 2†	0.56	0.23	1.40	0.22
Reticular radiologist 1†	1.10	0.77	1.57	0.62
Reticular radiologist 2†	2.56	1.18	5.58	0.018
Change radiologist 1				
Improved/stable	Ref			
Worse	0.91	0.08	10.03	0.94
Change radiologist 2				
Improved/stable	Ref			

Worse	1.65	0.15	18.74	0.69

† HR's are for a standard deviation change

Analysis adjusts for gender, pack years, percent predicted FVC at time2, percent predicted DLCO at time 2, and time between HRCTs.

Ref=reference