Full Title: Radiomic Profiling of Chest CT in a Cohort of Sarcoidosis Cases

Authors: Nichole E Carlson, PhD¹, William Lippitt, MS, PhD¹, Sarah M Ryan, PhD¹, Margaret Mroz, MS², Briana Barkes, MPH², Shu-Yi Liao, MD, MPH, ScD²,³,⁴, Lisa A Maier, MD²,³,⁴*, Tasha E Fingerlin, PhD¹,⁵*

*These authors contributed equally to this work.

Institutions: ¹Colorado School of Public Health, Department of Biostatistics and Informatics; ²National Jewish Health, Department of Medicine; ³University of Colorado, Department of Medicine; ⁴Colorado School of Public Health, Department of Environmental and Occupational Health; ⁵National Jewish Health, Department of Immunology and Genomic Medicine;

Corresponding Author:
Nichole E Carlson, PhD
Professor
Department of Biostatistics and Informatics
University of Colorado- Anschutz Medical Campus
13001 E 17th PL, MS B-119
Aurora, CO 80045
Tel: 303-724-4354
Fax: 303-724-4491
e-mail: nichole.carlson@cuanschutz.edu

Running Head: Radiomic Profiling of Sarcoidosis of the Lung
Supplementary Material Statement: This article has an online supplement, which is accessible from this issue’s table of contents online.

Summary of COI:
NEC, WLL, MM, TEF, SL, BB, TEF have received grants from the National Institutes of Health (R01 HL114587; R01 HL142049; U01 HL112695; U01 HL112707; U01 HL112707, U01 HL112694, U01 HL112695, U01 HL112696, U01 HL112702, U01 HL112708, U01 HL112711, U01 HL112712)

WLL is additionally support on National Institutes of Health 5T32HL007085

LAM has received grants from the National Institute of Health (R01HL140357, R01HL142049, R01 HL114587; U01 HL112695; U01 HL112707; U01 HL112707, U01 HL112694, U01 HL112695, U01 HL112696, U01 HL112702, U01 HL112708, U01 HL112711, U01 HL112712, and R01HL136681), Ann Theodore Foundation, the FSR, Mallinckrodt Pharmaceuticals, and the University of Cincinnati (Mallinckrodt Pharmaceuticals Foundation Grant) and serves on the Scientific Advisory Board for FSR, and Boeringer Ingelheim.

Funding: This work is supported by NIH grants: R01 HL114587; R01 HL142049; U01 HL112695; U01 HL112707; U01 HL112707, U01 HL112694, U01 HL112695, U01 HL112696, U01 HL112702, U01 HL112708, U01 HL112711, U01 HL112712 and 5T32HL007085

Acknowledgements and Author contributions: NEC is responsible for all content in the manuscript.

N.E.C. Substantial contributions to the conception or design of the work, lead biostatistical analysis and interpretation of the data and wrote manuscript. W.L. developed the filtering method for this analysis, performed the sensitivity and validation analyses and edited the manuscript. S.M.R. performed image processing, conducted
statistical analysis, prepared figures and table, interpreted the data, and contributed to the manuscript. M.M. coordinated the study, cleaned data, and contributed to the manuscript. B.B. supported data acquisition, data cleaning, project management, and contributed to the manuscript. S.L. contributed to the scientific rationale and interpretation of the data and contributed to the manuscript. L.A.M. Substantial contributions to the conception or design of the work, interpretation of the data and wrote the manuscript. T.E.F. Substantial contributions to the conception or design of the work, interpretation of the data and wrote the manuscript.

This work was supported by the National Institutes of Health (R01 HL114587; R01 HL142049; U01 HL112695; T32HL007085). Data from the GRADS study was used, which was funded by the NIH grant U01 HL112707 entitled “Sarcoidosis and A1AT Genomics and Informatics Centre”, as well as others (U01 HL112707, U01 HL112694, U01 HL112695, U01 HL112696, U01 HL112702, U01 HL112708, U01 HL112711, U01 HL112712). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Heart, Lung, and Blood Institute or the National Institutes of Health.
Abstract

Background: High resolution computed tomography (HRCT) of the chest is increasingly used in clinical practice for sarcoidosis. Visual assessment of chest HRCTs in patients with sarcoidosis has high inter- and intra-rater variation. Radiomics offers a reproducible quantitative assessment of HRCT lung parenchyma and could be useful as an additional summary measure of disease. We develop radiomic profiles on HRCT and map them to radiologic, clinical, and patient reported outcomes.
Research Question: Can radiomic analysis of chest HRCT cluster patients into groups that are related to radiologic, clinical, and patient reported outcomes?

Study Design and Methods: Three-dimensional radiomic features were calculated on chest HRCT for both lungs from sarcoidosis cases enrolled in the Genomic Research in Alpha-1 Antitrypsin Deficiency and Sarcoidosis (GRADS) study (N=320). Robust and sparse K-means was used to cluster sarcoidosis cases using their radiomic profiles. Differences in patterns on visual assessment (VAS) by cluster were identified using chi-squared tests. Linear regression investigated how pulmonary function tests and patient reported outcomes differed between clusters with and without adjustment for other radiologic quantification.

Results: Radiomic-based clustering identified four clusters associated with both Scadding stage and Oberstein score ($P<0.001$). One of the clusters had markedly few abnormalities. Another cluster had consistently more abnormalities along with more Scadding stage IV. Average pulmonary function testing (PFT) differed between clusters, even after accounting for Scadding stage and Oberstein score ($P <0.001$), with one cluster having more obstructive disease. The most discriminative radiomic measures explained 10-15% of the variation in PFT beyond demographic variables. Shortness of breath, fatigue, and physical health differed by cluster ($P <0.014$).

Interpretation: Radiomic quantification of sarcoidosis identifies new subtypes representative of existing radiologic assessment and more predictive of pulmonary function. These findings provide evidence that radiomics may be useful for identifying new imaging-based disease phenotypes.
Abstract word count: 295
Sarcoidosis is a granulomatous interstitial lung disease which affects ~ 110 thousand individuals in the United States. Typical diagnosis is between 30-50 years of life, resulting in decreases in quality of life (QOL) and productivity. Pulmonary disease occurs in over 90% of those with sarcoidosis with significant morbidity and mortality. Currently, visual assessment of chest radiography (CXR) is used to quantify lung abnormalities standardized via the Scadding staging. Substantial variation exists in CT-based radiographic patterns within each Scadding stage limiting its utility in predicting prognosis even in the extreme stages.

Chest high resolution computed tomography (HRCT) is increasingly used in clinical practice to monitor disease as it offers more detailed visualization of parenchymal abnormalities (PA) compared to CXR. As with CXR, visual assessment of chest HRCT is used to evaluate abnormalities although there are limited standardized scoring metrics. This is due in part to the diverse and heterogeneous patterns present on chest HRCT in sarcoidosis, often with multiple patterns noted. These complexities result in high inter- and intra-rater variation. Recently, a Delphi study was undertaken to define phenotypes based on visual assessment, yet without assessment of clinical utility. As not all patients have access to expert visual interpretation, more automated systems that quantify sarcoidosis chest HRCT could decrease variation and make HRCT more usable in routine clinical care.

Radiomics is when large numbers of quantitative features are extracted from medical images. A radiomics panel computes summary measures of the distribution of the...
Hounsfield units (HU) along with summary measures of the spatial relationships of neighboring voxels. The result is a characterization of image texture.

Radiomics have proved useful for quantifying HRCT in emphysema, idiopathic pulmonary fibrosis, interstitial lung disease, diffuse lung disease and cancer. Ryan et al. showed the potential utility of radiomics in sarcoidosis, comparing radiomic measures between sarcoidosis patients and controls. It remains unclear whether radiomics also has the potential to differentiate varied phenotypes within sarcoidosis patients and how radiomics relates to visual assessment (VAS). Radiomic profiles within patients with sarcoidosis that correlate with VAS, pulmonary function testing (PFT), and patient reported outcomes (PRO) would indicate that radiomics may serve as a useful refined measure to track change in the lung parenchyma over time than is possible with visual assessment.

The goal of this study was to develop a radiomic profile of chest HRCT in sarcoidosis using a phenotypically-diverse population of sarcoidosis participants from the Genomic Research in Alpha-1 Antitrypsin Deficiency and Sarcoidosis (GRADS) study. We employ statistical clustering techniques and investigate the clinical utility of the clusters by quantifying their association with VAS, PFT and PRO.

**Study Design and Methods**

**Study design and participants**
The sarcoidosis population was recruited in the multicenter NHLBI-funded GRADS study\textsuperscript{22} as cross-sectional observational cohort (N=368). This ancillary study has GRADS approval and all participants provided informed consent (IRB approval HS-2779 and HS-2780). More details of this cohort can be found in e-Appendix 1.

To comply with image biomarker standardization initiative (IBSI) recommendations\textsuperscript{25} details of the HRCT acquisition, processing and segmentation can be found in e-Appendix 1 and e-Tables 1 and 2. GRADS’ visual assessment score (VAS; Table 3, e-Appendix 1, e-Tables 3 and 4) was used to quantify overall Oberstein score\textsuperscript{5,22} along with additional information regarding presence of lymphadenopathy (LN), airway and vasculature distortion (AD), and PA (Table 3). Scadding stage was evaluated at the site using CXR (e-Appendix1).

PFT included pre-bronchodilator (pre-BD) forced expiratory volume at one second (FEV1), forced vital capacity (FVC), the ratio of FEV1 to FVC, and post-BD diffusing capacity of the lungs for carbon monoxide (DLCO). The PROs included the gastroesophageal reflux disease questionnaire (GERDQ)\textsuperscript{26}, the University of California San Diego Shortness of Breath Questionnaire (SOBQ)\textsuperscript{27}, two measures of fatigue (the Fatigue Assessment Scale [FAS]\textsuperscript{28} and Patient-Reported Outcomes Measurement Information System fatigue profile [PROMIS]\textsuperscript{29}), the Cognitive Failure Questionnaire (CFQ)\textsuperscript{30}, and the SF-12\textsuperscript{31,32} physical and mental subscales. The final analysis dataset included N=320 patients who had an analyzable HRCT and clinical data (e-Figure 1).
Radiomic Analysis

We computed 44 first-order and 239 gray-level co-occurrence matrix (GLCM) radiomic features on each lung (566 features) using the lungct and RIA packages in R\textsuperscript{33–36}. These are open-source packages with published documentation and permit perfect reproducibility with transparent implementation, aligning with IBSI goals. Furthermore, this allowed a fully R based analysis pipeline and provided a more comprehensive radiomic panel. Radiomic features beyond those typically used in neuroscience and cancer and previously standardized by IBSI may be important in diffuse pulmonary disease. For consistency with package documentation, we use the nomenclature of the RIA package.

To calculate the gray level co-occurrence matrix (GLCM) features, the Hounsfield units from each HRCT were discretized into 16 bins with equal relative frequencies; then, the features were calculated in 13 directions, assuming a voxel distance of one; these features were summarized using the mean statistic\textsuperscript{34,35}.

The R ez.combat function harmonized radiomic measures across scanners\textsuperscript{37–39}, while preserving the biological variability in age, height, BMI, sex, race-ethnicity and GRADS phenotype (e-Table 2).

Radiomic features were high-dimensional and repetitive (Figure 1). We used a decorrelation filter\textsuperscript{40} (e-Appendix 1) that prioritized first-order over second-order features.
to identify a representative feature subset for analysis. This reduced the features to 99.

We used robust and sparse k-means\textsuperscript{41} to cluster participants (R package, RSKC). For outlier robustness, trimming was set at 0.1. The optimal bound on feature weights (9.5) and number of clusters (4) were simultaneously selected using a permutation approach and BCS-based Gap statistic\textsuperscript{41} after standardization\textsuperscript{42}. Standardization was performed with respect to the permutation reference to permit comparability across bounds and cluster numbers. The top five discriminative features were selected for further investigation.

Statistical Analysis and Validation

Descriptive statistics on VAS were computed by cluster and simulated Fisher’s exact tests used to assess cluster differences. Linear regression quantified associations between cluster, Scadding stage, and Oberstein Score, and each of the outcomes. We fitted univariable models with cluster, Scadding stage, or Oberstein score and then modeled them together to determine if cluster association remained significant accounting for other radiologic findings. Each outcome was modelled separately using a complete case analysis. Linear regression quantified associations between discriminative features and outcomes. All models were adjusted for age, sex, race/ethnicity (e-Appendix 1), height, and BMI. Analyses were conducted in R\textsuperscript{36}.

This work was an unsupervised problem, which made traditional training and test validation approaches difficult given unknown true groupings. Instead, we conducted internal validation by applying our analysis pipeline under various conditions and with bootstrapped samples (e-Appendix 1).
Results

Tables 1-2 show the characteristics of the study population (N=320), 146 (46%) self-identified as male, 220 (68.8%) self-identified as non-Hispanic white, 76 (24%) as black, 15 (4.7%) as Hispanic, and 9 (2.8%) as one of Asian, American Indian, Alaska Native, not identifying a single primary race or missing (N=2). The average age was 53 (SD=10) years, average height 67.0 in (SD=4.0) and average BMI 30.6 kg/m$^2$ (SD=6.5). By design participants spread across Scadding stages, with 43 (14%) stage 0, 63 (20%) stage 1, 92 (29%) stage 2, 44 (14%) stage 3 and 75 (24%) stage 4. The average FEV1 was 2.62 L (SD=0.89), FVC 3.57 L (SD=1.07), and DLCO 80.31 (SD=23.97). The population predominantly demonstrated non-obstructive FEV1/FVC ratio (N=235; 73%).

Radiomic Based Clustering

We identified four clusters (Table 1). Ordered from highest FVC to lowest, 56 (17.5%) patients were cluster 1, 110 (34.3%) cluster 2, 54 (16.9%) cluster 3, and 100 (31.3%) cluster 4. Cluster was associated with Scadding stage ($P<0.001$), but not a direct reflection of Scadding stage (Table 1). Descriptively, cluster 3 had the highest percentage of Scadding stage IV (48%). Clusters 1 and 2 each had approximately 60% of the patients with a Scadding stage of I or II, a higher proportion of stage 0, and limited Scadding stage IV present. Cluster 3 had the second highest observed Scadding stage IV (40%) with a similar percentage of Scadding stage II compared to clusters 1 and 2 (~30%).

Radiomic clustering and VAS
When descriptively investigating VAS patterns, the clusters reflected increased presence of VAS abnormalities with cluster 1 having low presence of nearly all airway and vascular distortions (AD) and PA and much lower average Oberstein scores than clusters 2-4 (Table 3; \( P < 0.001 \)). Cluster 2 reflected some increase in AD and PA with clusters 3 and 4 having a higher presence of AD and PA. For example, cluster 4 had at least double the presence of AD and PA compared to cluster 2. After adjustment for demographics, average Oberstein score for cluster 3 was 1.98 (SE=0.48) units and cluster 4 was 3.15 (SE=0.39) units higher than cluster 2 (\( p<0.001 \)).

**Associations with PFT and PRO**

After adjustment for demographics, average PFT differed between clusters (Figure 2; \( P<0.001 \)). Clusters 2 and 3 had \(~0.3\) L lower average FVC compared to clusters 1 and 4 had \(~0.9\) L lower average FVC compared to cluster 1. Cluster 2 had \(~0.2\) L lower average FVC compared to cluster 1 and cluster 3 had a larger decline with an average FEV1 \(~0.5\) L lower than cluster 1. Cluster 4 had the lowest average FEV1 and \(~0.8\) L lower than cluster 1. DLCO had a similar pattern to FVC. Cluster 3 had the most FEV1/FVC based obstruction (54%) with clusters 1 and 2 having the same amount of obstruction (16%) and cluster 4 falling in the middle (32%).

After adjustment for demographics, PROs also differed between clusters but not consistently (Figure 3). Fatigue and Average shortness of breath (SOBQ) differed between clusters (\( P<0.014 \)). Clusters 2 and 4 had a fatigue score that was more than 3...
units lower on average compared to cluster 1 (p<0.016). Cluster 4 also had an average
SOBQ ~12 units higher than cluster 1 (p=0.003).

The clusters remained significantly associated with PFT after adjusting for Scadding
stage (Figure 2; P<0.001) or Oberstein score (Figure 2; P<0.001) and after adjusting for
both simultaneously (Figure 2; P<0.001). Oberstein score was not significantly
associated with FVC after adjustment for cluster (P=0.59) and became even less
significant after further adjustment for Scadding stage (P=0.88). However, Scadding
stage was significantly associated with FVC after adjustment for cluster and Oberstein
score (P=0.019). The findings for Oberstein score with DLCO were like those of FVC.

For FEV1, Oberstein score was still significantly associated after adjustment for cluster
(P=0.011) but became insignificant after additional adjustment for Scadding stage
(P=0.21). Scadding stage remained significant in all models (P<0.001). For FEV1/FVC,
all three assessments (radiomics, Scadding, and Oberstein score) were significant
(P<0.016).

Cluster also remained associated with fatigue and SOBQ score after adjustment for
Scadding and Oberstein (P<0.038; Figure 3).

The five most discriminatory radiomic features included kurtosis, which is a measure of
shape of the distribution of the HU from an image, as well as four summary measures
from the gray level co-occurrence matrix (GLCM). The GLCM measure spatial
correlation and similarity of the HU in image voxels near each other. Image visualization
of cases with different values of kurtosis and two GLCM measures are shown in Figure 4 with distribution of HU in e-Figure 2 for maximum, median and minimum kurtosis levels in our population. The images visibly show the increase in observable PA (Figure 4) with lower kurtosis. In addition, Figure 5 shows how Oberstein Score separates based on kurtosis and GLCM Min. GLCM Min appears to be most related to severity with higher Oberstein score mapping with higher GLCM Min and Scadding stage is, in general, not associated with either measure.

The discriminatory radiomic measures were jointly associated with FVC, FEV1, FEV1/FVC and DLCO ($P<0.001$; Table 4 and e-Table 5 for PRO). The radiomic measures explained between 10 to 15% more variation in PFT than adjustment for demographics (age, race, sex, height and BMI) only. For comparison, Scadding stage explained between 4% and 12% more variation in PFT and Oberstein score explained between 2% and 8% more variation in PFT.

Validation of Clustering

Detailed results can be found in e-Appendix 2, e-Figure 3. The validation analysis suggested only mild sensitivity of clustering to the analysis pipeline. There is more sensitivity to bootstrapped samples in how the sample is divided into clusters; however, cluster consistently remained a significant predictor of FVC across all sources of random perturbations in the pipeline. In all validation cases the proportions of $P$-values less than 0.01 were 100%. 
Discussion

Radiographic manifestations in sarcoidosis are protean. As a result, traditionally VAS is used for image characterization and not standardized. The utility of VAS in sarcoidosis is limited by the intra-observer and inter-observer variability. Radiomics is a more reproducible and computationally efficient approach to characterize HRCT. We used radiomics to characterize images from a large, phenotypically diverse cohort of sarcoidosis subjects and demonstrated that radiomics are associated with VAS, clinical and patient reported outcomes of disease. Using a common unsupervised learning approach, we identified four radiomic based clusters. These clusters differed significantly by PFT, fatigue, and shortness of breath, two PROs. Notably, each cluster included a range of Scadding stages and cluster remained significantly associated with PFT after accounting for Scadding stage and Oberstein score. These data suggest that radiomics represents radiographic abnormalities that differ from Scadding stage and may be a better representation of image characterization to explain PFT and PRO.

A holistic assessment of the clusters showed two lower severity and two higher severity clusters (Table 5). One high severity cluster (4) had significant fibrosis, substantial presence of AD and PA on VAS, and decreased PFT, which resulted in more shortness of breath and fatigue. Cluster 3 represented those with obstructive spirometry and more presence of AD and PA than clusters 1 and 2, moderate fibrosis, and the lowest FEV1/FVC ratio. These interpretations have similarities to another recent attempt to subtype using clinical characteristics. Unlike our work, that work had no radiomics or VAS. We both several low severity clusters with minimal PA. We both found an
obstructive cluster and a severe fibrotic cluster. More recently, HRCT VAS phenotypes were proposed via a Delphi consensus study\textsuperscript{12}. That work defined two major groups, those with fibrotic versus non-fibrotic findings, similar to our clusters with fibrotic (3 and 4) versus more non-fibrotic (1 and 2) findings. However, the VAS of their various subtypes of non-fibrotic and fibrotic CT findings were included in a number of our clusters. This may reflect the power in our sample size to define subtypes or possibly that these phenotypes or VAS overlap as noted in our study.

Demographic characteristics of the individual explained a sizable amount of the variation observed in PFT (\(R^2=50-60\%\)). Radiomics also explained a significant amount of additional variation in PFT (10-50\%). This additional amount of variation explained is consistent with other quantitative imaging approaches such as CALIPER\textsuperscript{17} and those used for investigations of other lung conditions such as systemic sclerosis\textsuperscript{44} and diffuse interstitial lung diseases\textsuperscript{45}.

Except in lung cancer research, much of the quantitative CT image analysis has focused on HU density\textsuperscript{18,44}. We included summarization of both density and spatial characterization. More measures of GLCM appeared discriminative in the cluster analysis than densitometry measures. This implies that texture is perhaps more useful for differentiating PA in sarcoidosis.

Although it remains speculative as to why decreased kurtosis is associated with decreased FVC and DLCO, it is plausible that as kurtosis decreases, more severe PA
are present that impact lung function. As kurtosis decreases, more severe PA are visual
present that may impact the function of the lung (Figure 4). Severe PA in sarcoidosis
are observed as higher HU values. In e-Figure we observed the presence of a higher
frequency of higher HU values dampening the peak in the HU distribution and leading to
a lower kurtosis. In addition, GLCM Min appears likely to measure disease severity
based on the visual patterns with Oberstein in Figure 5.

The associations with cluster and PROs were less consistent, although important
patterns emerged. Cluster 4 demonstrated decreased QOL (SF-12 physical) and worse
shortness of breath (SOBQ), findings that are consistent with worse lung function and
VAS scores in this cluster. Cluster explained substantially less variation in PROs (0 to
8%) compared to PFTs. While the tools we used to measure PROs are validated and
used widely in sarcoidosis, they are not generally correlated with objective measures of
lung function. As an example, fatigue is a known multi-factorial PRO in sarcoidosis
and may impact a number of other PRO. Our finding that cluster is associated with
decreased physical function and increased shortness of breath provides
encouragement that more detailed characterization of lung abnormalities, like we
developed here, could contribute to a better understanding of the current disconnect
between measures of lung function and patient experiences.

This study had several strengths. To our knowledge, this was the largest cluster of
sarcoidosis patients with research grade HRCT available allowing for a full 3-D based
quantitative analysis. Radiomic and other quantitative imaging approaches have a major
strength in that they can be computed in a time efficient and reliable, reproducible automated procedure. The radiomics package in R took only 3 minutes to compute all the 3-D radiomics measures. This means large quantities of scans can have radiomic profiles computed for aiding in decisions on what scans the radiologist might prioritize for further consideration for visual or clinician evaluation.

In addition to the high-quality and robust clinical data, the GRADS study also provides a comprehensive and consistently collected set of patient reported outcomes, which reflect important aspects of treatment decision-making processes. The results internally validated suggest our approach is not over fitting these data.

This study is not without limitations. The GRADS study relied on enrollment from academic centers and could be skewed to a population that was referred for worse disease or was near one of the centers, which were primarily localized to the Eastern US. The demographic is predominantly white and of higher SES as a result and thus may not provide full representation of the spectrum of disease. In addition, many subjects had disease for a decade or longer and were treated. While the same protocol was used across study sites for obtaining the CT images we studied, the scanners themselves differed. Harmonization was performed to mitigate systematic differences due to scanner type. In addition, the distribution of PFT measures is not dependent on scanner type. As such, we expect any differences in the radiomic measures due to scanner type to be non-differential as they relate to PFTs. Finally, the number of clusters we identified may not directly translate to other populations of sarcoidosis or be
the unique optimal solution. There are multiple ways to choose the optimal number of clusters. We used an alternative radiomics package with a broader set of measures. However, not all these measures were part of IBSI work. We used open-source software accessible to any reader for verification of our findings and to conduct comparisons with other software.

Interpretation

In summary, this work provides evidence that in sarcoidosis radiomic quantification is useful for classifying abnormality of the lung along with pulmonary function and to a lesser degree patient reported outcomes. Future work should evaluate the potential of radiomics to capture small changes over time not easily detected by other assessment to further evaluate the potential of radiomics serving as a clinically useful quantitative measure of sarcoidosis presentation and progression.
References


32. Ware J, Kosinski M, Turner-Bowker D, Gandek B. How to score SF-12 items. *SF-12 V2 Score Version 2 SF-12 Health Surv* 2002;29–38.


40. Lippit WL. Clustering with Highly Correlated Features. *MS Thesis Univ Colo* 2022;


Take-Home Points

Study Question: Is radiomics a useful quantitative approach for defining radiographic subtypes in sarcoidosis?

Results: Radiomics find four radiographic subtypes in sarcoidosis that are related to visual assessment and more predictive of clinical measures such as pulmonary function that visual assessment.

Interpretation: Radiomics analysis suggests four subtypes of sarcoidosis ranging from low severity with limited radiographic abnormality and normal pulmonary function to severe with significant abnormality on chest CT and either obstruction based on pulmonary function or reduced pulmonary function with fibrotic presentation on image.
Table 1: Patient demographics by radiomic cluster, ordered from least severe (1) to most severe (4) based on average FVC. Unless otherwise noted values are mean (SD).

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>N</th>
<th>Overall</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample Size</td>
<td>320</td>
<td>56</td>
<td>110</td>
<td>54</td>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FVC (L)</td>
<td>6</td>
<td>3.57 (1.14)</td>
<td>3.91 (1.16)</td>
<td>3.88 (1.08)</td>
<td>3.61 (1.27)</td>
<td>3.02 (0.92)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Age (yr)</td>
<td>0</td>
<td>52.9 (9.9)</td>
<td>49.8 (10.6)</td>
<td>53.1 (9.9)</td>
<td>52.5 (9.8)</td>
<td>54.8 (9.3)</td>
<td>0.026</td>
</tr>
<tr>
<td>Female; N (%)</td>
<td>174</td>
<td>37 (66%)</td>
<td>52 (47%)</td>
<td>36 (67%)</td>
<td>49 (49%)</td>
<td></td>
<td>0.021</td>
</tr>
<tr>
<td>Race/Ethnicity; N (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.033</td>
</tr>
<tr>
<td>Asian, American</td>
<td>9</td>
<td>(2.8%)</td>
<td>0 (0%)</td>
<td>4 (3.6%)</td>
<td>1 (1.9%)</td>
<td>4 (4.0%)</td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>76</td>
<td>(24%)</td>
<td>12 (21%)</td>
<td>16 (15%)</td>
<td>13 (24%)</td>
<td>35 (35%)</td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td>15</td>
<td>(4.7%)</td>
<td>2 (3.6%)</td>
<td>6 (5.5%)</td>
<td>1 (1.9%)</td>
<td>6 (6.0%)</td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>0</td>
<td>(0%)</td>
<td>42 (75%)</td>
<td>84 (76%)</td>
<td>39 (72%)</td>
<td>55 (55%)</td>
<td></td>
</tr>
<tr>
<td>Height (in)</td>
<td>0</td>
<td>67.0 (4.0)</td>
<td>66.3 (4.2)</td>
<td>67.7 (4.2)</td>
<td>66.5 (3.7)</td>
<td>66.9 (3.7)</td>
<td>0.13</td>
</tr>
<tr>
<td>BMI (kg/m^2)</td>
<td>0</td>
<td>30.6 (6.5)</td>
<td>30.2 (7.4)</td>
<td>32.6 (5.7)</td>
<td>27.0 (5.8)</td>
<td>30.7 (6.4)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Scadding; N (%)</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>II</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>III</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IV</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

CC-BY-NC-ND 4.0 International license. It is made available under a CC-BY-NC-ND 4.0 International license.
Table 2: Summary measures of pulmonary function testing and self-reported outcomes by radiomic cluster ordered from least severe (1) to most severe (4) based on average FVC. Unless otherwise noted values are mean (SD).

<table>
<thead>
<tr>
<th></th>
<th>N Miss</th>
<th>Overall</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample Size</td>
<td>320</td>
<td>56</td>
<td>110</td>
<td>54</td>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FEV1 (L)</td>
<td>6</td>
<td>2.62 (0.05)</td>
<td>2.98 (0.13)</td>
<td>2.96 (0.08)</td>
<td>2.44 (0.15)</td>
<td>2.14 (0.07)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>FVC (L)</td>
<td>6</td>
<td>3.57 (0.06)</td>
<td>3.91 (0.16)</td>
<td>3.88 (0.10)</td>
<td>3.61 (0.18)</td>
<td>3.02 (0.09)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>FEV1/FVC (%)</td>
<td>6</td>
<td>0.73 (0.01)</td>
<td>0.76 (0.01)</td>
<td>0.77 (0.01)</td>
<td>0.67 (0.02)</td>
<td>0.71 (0.01)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Obstructive Type&lt;sup&gt;a&lt;/sup&gt;; N (%)</td>
<td>6</td>
<td>85 (27%)</td>
<td>9 (16%)</td>
<td>17 (16%)</td>
<td>27 (54%)</td>
<td>32 (32%)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>DLCO (mL/min/mmHg)</td>
<td>18</td>
<td>22.07 (0.45)</td>
<td>24.13 (1.03)</td>
<td>25.33 (0.66)</td>
<td>21.30 (1.01)</td>
<td>17.56 (0.76)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>FAS&lt;sup&gt;a&lt;/sup&gt;</td>
<td>140</td>
<td>29.77 (0.41)</td>
<td>32.46 (1.29)</td>
<td>28.95 (0.66)</td>
<td>30.06 (0.91)</td>
<td>29.24 (0.65)</td>
<td>0.038</td>
</tr>
<tr>
<td>GERDQ&lt;sup&gt;b&lt;/sup&gt;</td>
<td>1</td>
<td>7.07 (0.12)</td>
<td>7.04 (0.30)</td>
<td>7.09 (0.17)</td>
<td>6.72 (0.27)</td>
<td>7.25 (0.23)</td>
<td>0.5</td>
</tr>
<tr>
<td>CFQ&lt;sup&gt;c&lt;/sup&gt;</td>
<td>3</td>
<td>32.79 (0.95)</td>
<td>34.57 (1.87)</td>
<td>34.14 (1.58)</td>
<td>32.50 (2.22)</td>
<td>30.44 (1.93)</td>
<td>0.4</td>
</tr>
<tr>
<td>SOBQ&lt;sup&gt;d&lt;/sup&gt;</td>
<td>17</td>
<td>26.28 (1.31)</td>
<td>22.08 (3.13)</td>
<td>19.94 (1.80)</td>
<td>26.98 (3.12)</td>
<td>35.04 (2.56)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>PROMIS Fatigue&lt;sup&gt;e&lt;/sup&gt;</td>
<td>38</td>
<td>25.51 (0.52)</td>
<td>25.82 (1.28)</td>
<td>23.96 (0.82)</td>
<td>26.80 (1.35)</td>
<td>26.18 (0.93)</td>
<td>0.2</td>
</tr>
<tr>
<td>SF12-Physical&lt;sup&gt;f&lt;/sup&gt;</td>
<td>6</td>
<td>41.70 (0.63)</td>
<td>44.78 (1.55)</td>
<td>42.23 (1.10)</td>
<td>42.80 (1.49)</td>
<td>38.77 (1.07)</td>
<td>0.009</td>
</tr>
<tr>
<td>SF12-Mental&lt;sup&gt;h&lt;/sup&gt;</td>
<td>6</td>
<td>48.12 (0.56)</td>
<td>47.13 (1.21)</td>
<td>48.60 (0.90)</td>
<td>47.33 (1.49)</td>
<td>48.56 (1.09)</td>
<td>0.7</td>
</tr>
</tbody>
</table>

<sup>a</sup>FAS range 5 – 50, higher values indicate more fatigue; <sup>b</sup>GERDQ range 0 – 18, higher values indicate a higher likelihood of GERD; <sup>c</sup>CFQ range 0 -100, higher indicates more cognitive failure; <sup>d</sup>SOBQ range 0 – 120, higher values indicate more shortness of breath; <sup>e</sup>PROMIS Fatigue range 10-50, higher values indicate more fatigue; <sup>f</sup>SF12-Physical-For general U.S.
population, mean = 50, SD = 10 and higher values indicate better physical QOL; SF12-Mental-
For general U.S. population, mean = 50, SD = 10 and higher values indicate better mental QOL.
Table 3: Distribution of VAS measures by radiomic cluster. Three participants were missing VAS and Scadding stage. All measures represent the abnormality was present (vs. absent) and are presented as N (%) unless otherwise noted. LN = lymphadenopathy; BVB = Bronchovascular Bundle

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Overall</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample Size</td>
<td>317</td>
<td>56</td>
<td>108</td>
<td>53</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>Mediastinal LN</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>172 (54%)</td>
<td>19 (34%)</td>
<td>60 (56%)</td>
<td>29 (55%)</td>
<td>64 (64%)</td>
<td>0.003</td>
</tr>
<tr>
<td>Hilar LN</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>135 (43%)</td>
<td>14 (25%)</td>
<td>46 (43%)</td>
<td>26 (49%)</td>
<td>49 (49%)</td>
<td>0.017</td>
</tr>
<tr>
<td>Micronodule</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>147 (46%)</td>
<td>23 (41%)</td>
<td>46 (43%)</td>
<td>29 (55%)</td>
<td>49 (49%)</td>
<td>0.4</td>
</tr>
<tr>
<td>Airway and Vascular Distortion (AD)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BVB Distortion</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>202 (64%)</td>
<td>19 (34%)</td>
<td>53 (49%)</td>
<td>43 (81%)</td>
<td>87 (87%)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Traction Bronchiectasis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>129 (41%)</td>
<td>6 (11%)</td>
<td>25 (23%)</td>
<td>32 (60%)</td>
<td>66 (66%)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Parenchymal Opacity and Distortion (PD)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ground Glass</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>110 (35%)</td>
<td>5 (8.9%)</td>
<td>22 (20%)</td>
<td>21 (40%)</td>
<td>62 (62%)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Honeycombing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>25 (7.9%)</td>
<td>0 (0%)</td>
<td>2 (1.9%)</td>
<td>4 (7.5%)</td>
<td>19 (19%)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Reticular Abnormality</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>88 (28%)</td>
<td>3 (5.4%)</td>
<td>17 (16%)</td>
<td>21 (40%)</td>
<td>47 (47%)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Mosaic Attenuitation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>86 (27%)</td>
<td>10 (18%)</td>
<td>18 (17%)</td>
<td>14 (26%)</td>
<td>44 (44%)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Interlobular Septal Thickening</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>54 (17%)</td>
<td>5 (8.9%)</td>
<td>14 (13%)</td>
<td>7 (13%)</td>
<td>28 (28%)</td>
<td>0.007</td>
</tr>
<tr>
<td>Oberstein Overall; Mean (SE)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4.23 (0.18)</td>
<td>2.11 (0.32)</td>
<td>2.95 (0.26)</td>
<td>5.32 (0.44)</td>
<td>6.23 (0.28)</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>
Table 4: Results of the multivariable regression analysis (coefficients and SE’s) of the five discriminatory radiomic measures for PFT. \(^d\)

<table>
<thead>
<tr>
<th>Feature</th>
<th>GLCM Gaussian</th>
<th>GLCM Inv Gaussian-E</th>
<th>Kurtosis GLCM</th>
<th>GLCM Sum Energy</th>
<th>GLCM Min</th>
<th>P-value</th>
<th>Rsq base</th>
<th>Rsq</th>
</tr>
</thead>
<tbody>
<tr>
<td>FVC</td>
<td>-0.02 (0.11)</td>
<td>-0.06 (0.08)</td>
<td>0.44 (0.08)(^a)</td>
<td>0.04 (0.09)</td>
<td>-0.06 (0.08)</td>
<td>&lt;0.0001</td>
<td>0.726</td>
<td>0.622</td>
</tr>
<tr>
<td>FEV1</td>
<td>-0.16 (0.10)</td>
<td>-0.17 (0.08)(^c)</td>
<td>0.08 (0.07)</td>
<td>0.01 (0.08)</td>
<td>0.23 (0.08)(^a)</td>
<td>&lt;0.0001</td>
<td>0.619</td>
<td>0.514</td>
</tr>
<tr>
<td>FEV1/FVC</td>
<td>-0.04 (0.02)(^b)</td>
<td>-0.03 (0.01)(^b)</td>
<td>-0.07 (0.01)(^b)</td>
<td>-0.02 (0.01)</td>
<td>0.09 (0.01)(^b)</td>
<td>&lt;0.0001</td>
<td>0.199</td>
<td>0.049</td>
</tr>
<tr>
<td>DLCO</td>
<td>-2.23 (0.85)(^a)</td>
<td>-1.65 (0.65)(^b)</td>
<td>1.40 (0.62)(^c)</td>
<td>0.50 (0.70)</td>
<td>-0.04 (0.68)</td>
<td>&lt;0.0001</td>
<td>0.636</td>
<td>0.541</td>
</tr>
</tbody>
</table>

\(^a\) P-value <0.001; \(^b\) P-value<0.025; \(^c\) P-value<0.035. \(^d\)Each linear regression model included all five radiomic features and was additionally adjusted for age, gender, race/ethnicity, BMI and height. Bolded cells are statistically significant (p-values in the footnote). GLCM Gaussian is the sum of the GLCM with a Gaussian weight applied. GLCM-Inv Gaussian-E is the sum of the GLCM for the inverse Gaussian weighting scheme. GLCM Sum Energy is the sum of the squared values in the GLCM and a measure of uniformity. GLCM Min is the minimum of the GLCM. P-value is for the partial F-test of significance of any of the five radiomic measures.
Table 5: Holistic interpretation of radiomic clusters in terms of VAS, PFT, and PRO measures.

<table>
<thead>
<tr>
<th>Cluster #</th>
<th>Cluster Description</th>
<th>Directional changes in characteristics</th>
<th>Severe Disease Features$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Low severity with less LN, less PA and AD and limited shortness of breath (PRO)</td>
<td>Oberstein Score↓ Scadding stage 1,2 FEV1, FVC, DLCO↑ LN↓ vs. Cluster 3,4 SOBQ↓ vs. Cluster 4 Age↓ &amp; Female↑</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Low severity with more LN, smaller increases in PA and AD and limited PFT and shortness of breath</td>
<td>Oberstein Score↓ Scadding stage 1,2 FEV1, FVC, DLCO↑ LN↑ vs. Cluster 1 SOBQ↓ vs. Cluster 4 Age ↔, Balanced Sex</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Severe obstructive, fibrotic presentation with significant PA, AD, and decreased PFT</td>
<td>Oberstein Score↑ vs. 1,2↓ 4 Scadding stage 2,4 FEV1/FVC↓ (Obstruction↑), PFT↓ LN → vs. 2 &amp; 4 Age →, Female↑</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Severe PA and AD with significant fibrosis; low PFT and increased shortness of breath and fatigue.</td>
<td>Oberstein Score↑ Scadding stage 4 FEV1, FVC, DLCO↓ LN → vs. 2 &amp; 3 SOBQ, Fatigue↑ Age →, Balanced Sex</td>
<td></td>
</tr>
</tbody>
</table>

$^a$ No fill represents similar to least impairment observed. Lined shading represents moderate impairment or ~25-50% prevalence for Scadding stage. Dark shading represents similar to worst impairment observed or ~50% prevalence for Scadding stage.
Figure Legends

Figure 1: Heat map of the correlation between different radiomic measures for the entire population.

Figure 2: Radiomic cluster differences in average PFT values (columns) from regression models with adjustment for demographics (top row), then additionally adjusted for either Scadding stage or Oberstein score (middle two rows), then adjusted for both Scadding stage and Oberstein score (bottom row). The significance of the cluster differences from cluster 1 are in green (p>0.05) and blue (p<0.05). The bottom text in each panel is the overall p-value for the association between clustering and PFT outcome along with the R².

Figure 3: Radiomic cluster differences in average PRO values (columns) from regression models with adjustment for demographics (top row), then additionally adjusted for either Scadding stage of Oberstein score (middle two rows), then adjusted for both Scadding stage and Oberstein score (bottom row). The significance of the cluster differences from cluster 1 are in green (p>0.05) and blue (p<0.05). The bottom text in each panel is the overall p-value for the association between clustering and PFT outcome along with the R².

Figure 4: CT images in axial orientation for three patients with minimum, median, and maximum values for GLCM Gaussian (left column), the GLCM Inverse Gaussian, and the kurtosis.

Figure 5: Separation of Oberstein score and Scadding based kurtosis (y-axis) and GLCM-Min (x-axis) two of the five most important radiomic variables in defining clusters. Colors represent the Oberstein score and Scadding stage.
<table>
<thead>
<tr>
<th>Max</th>
<th>Median</th>
<th>Min</th>
</tr>
</thead>
</table>

Kurtosis

[Images of lung CT scans]