Feature Analysis of Hyperpolarized Helium-3 Pulmonary MRI in Asthmatics versus Non-Asthmatics

N. J. Tustison[1], T. A. Altes[2], G. Song[1], J. P. Mugler III[2], E. E. de Lange[2], and J. C. Gee[1]

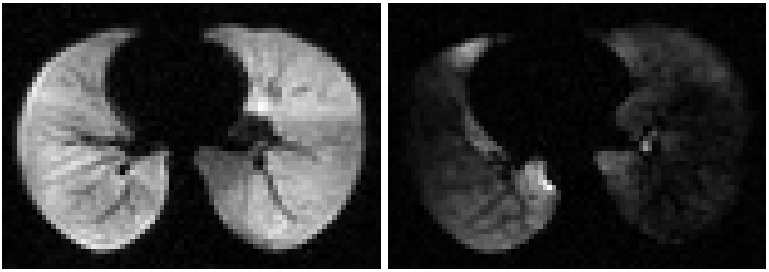
[1] Penn Image Computing and Science Laboratory, University of Pennsylvania, Philadelphia, PA, USA
[2] Department of Radiology, University of Virginia, Charlottesville, VA, USA

Abstract. Computational techniques for parenchymal characterization using CT have demonstrated significant research potential [1, 2]. We describe similar quantitative analysis of the lung using hyperpolarized helium-3 MR ventilation image data. This study consisted of a total of 55 subjects (47 asthmatics and 8 non-asthmatics). Each subject was imaged both before and after respiratory challenge. Additionally, each subject underwent a complete pulmonary function exam. Following image acquisition, approximately 1600 statistical features were calculated from the segmented lungs (and sub-regions) in each image. Each of these features were ranked along with the 27 pulmonary function test (PFT) values using a mutual information based feature subset selection algorithm. It is shown that several image features perform much better in characterizing clinical diagnosis compared with the current clinical gold-standard PFT values.

1 Introduction

Recent developments in MRI research utilizing noble gases, such as helium-3 and xenon-129, have demonstrated the capability of visualizing alveolar and bronchial air spaces [3]. Currently, hyperpolarized helium-3 MRI is a low-risk investigatory technique which provides high spatial and temporal resolution images of the air spaces of the lungs.

Ventilation or spin density images are acquired by measuring the signal produced by the helium-3 atoms within each voxel. If the signal intensity in the MR images were solely dependent upon the density of helium-3 atoms in each pixel, these images would directly reflect regional ventilation. However, this is not the case since the coil transmit and receive sensitivity, and the regional partial pressure of oxygen within the lung contribute to the measured signal intensity. For this reason, the helium-3 images provide information about the homogeneity of ventilation within the lung but do not provide a quantitative measure of absolute regional ventilation. When a subject inhales the helium-3 gas, areas of the lung that are well ventilated receive a large volume of helium-3 gas which produces a strong MR signal causing these areas to appear bright on the resulting MR images [3]. In contrast, regions that are poorly ventilated receive little helium-3 and thus produce a weak MR signal causing these regions to appear dark in the resulting images (Fig. 1).



(a) (b)

Fig. 1. Axial hyperpolarized helium-3 MR ventilation images contrasting well-ventilated lungs (left) with poorly ventilated lungs (right).

Previous methodologies for evaluation of helium-3 ventilation images have been limited to radiological assessment of signal heterogeneity as well as various forms of quantifying *ventilation defects*, i.e. regions of poor ventilation, whether it be the number of such defects or the total volume of such defects. Using hyperpolarized helium-3 MRI, it has been shown that asthmatics have an increased volume of lung regions that are poorly ventilated (ventilation defects) than age-matched normal subjects, and that these defects increase in number with increasing asthma severity or with provocation such as exercise or methacholine [4]. Thus, the ventilation defects on hyperpolarized gas MRI appear to be depicting the reversible airway obstruction that was known to occur in asthmatics but was previously difficult to visualize [5].

In this paper, we describe an automated computational framework for generating and analyzing features from helium-3 ventilation images. Such features attempt to characterize lung parenchyma in the varying stages of pathogenesis. Our computational approach permits the analysis of large studies due to complete automation, is extendible to incorporating new features, and facilitates quick processing of data. Furthermore, we demonstrate that our analysis compares favorably with the gold-standard of pulmonary function testing for characterizing clinical diagnosis of asthmatics.

2 Method

Image processing for each helium-3 image volume requires the following steps:

1. bias field correction,

2. segmentation of the lungs,
3. generation of regional masks,
4. generation of feature images,
5. calculation of the image features, and
6. ranking of the image features in conjunction with the PFT values.

Due to the large quantity of current image data used for the study in addition to future stuides, a software pipeline was developed for automated analysis of helium-3 lung MRI. This pipeline was implemented on a computing cluster that allows for distributed parallel processing of the images. In addition, much of the software development constituting our analysis pipeline has been made publicly available through the Insight Toolkit (ITK) of the National Institutes of Health, an open-source repository for popular image analysis algorithms [6].

2.1 Retrospective Bias Field Correction

Significant bias field effects can be seen in typical helium-3 images. This is manifested as a low frequency intensity artifact across the image. The left column of Figure 2 illustrates this bias field artifact in two subjects. Several algorithms exist for correcting the nonparametric nonuniform intensity in magnetic resonance images caused by field inhomogeneities. One popular algorithm is the non-uniform intensity normalization (N3) approach [7, 8]. A particularly advantageous aspect of this algorithm is that it does not require a prior tissue model for its application. This algorithm was used to estimate the bias field and correct the images shown respectively in the middle and right columns of Figure 2.

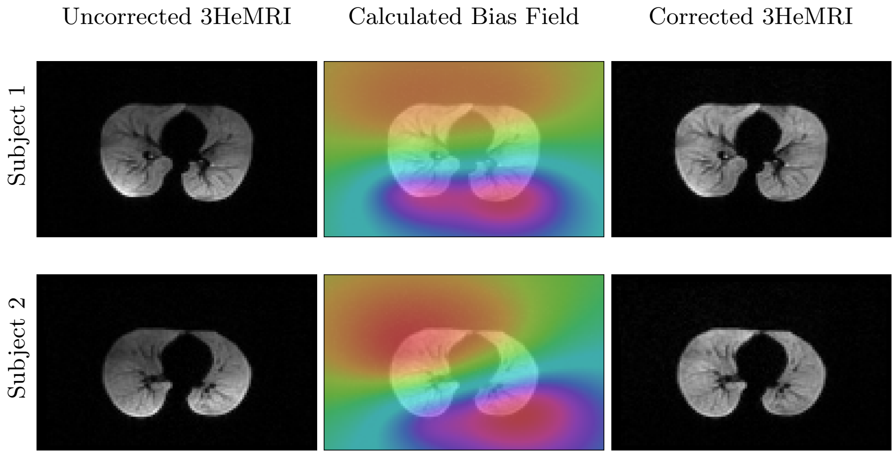
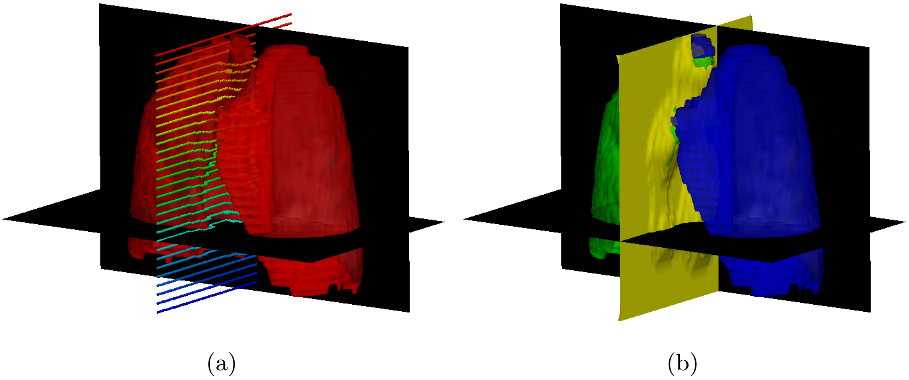


Fig. 2. Left column: Axial 3HeMRI from two subjects evidencing severe bias field artifacts. Middle column: The calculated bias field. Right column: Corrected images.

2.2 Lung Segmentation

An essential precursor to calculation of meaningful image measures is the segmentation of whole lungs from the helium-3 images. Because the quantity of data processed prohibits the use of routine manual segmentation or supervision-intensive semi-automated segmentation methods, we developed an automated segmentation routine specifically tailored for helium-3 lung images.

We first preprocess the images by applying an anisotropic diffusion operation which smoothes both the background noise and the helium-3 signal internal to the lungs while respecting the lung boundaries. Using the statistics surrounding a seed point placed in the background of the image, we grow this region to the rest of the connected background region while iteratively recalculating regional statistics. This separates the whole lung from the background. A 3-D view of a single segmentation is shown in Fig. 3(b).



(a) (b)

Fig. 3. Automatic labeling of left and right lungs. (a) Segmentation of the lungs from the background is illustrated in red. For each axial slice, we generate a minimum-cost path (the minimum path for each axial slice is rendered a different color). (b) Fitting a B-spline surface to these minimum-cost paths labels the left (blue) and right (green) lungs.

Due to the anatomical proximity of the anterior portions of the left and right lungs as well as the presence of helium-3 in the trachea, the whole lung segmentation might produce a single connected component segmented object. To separate the left and right lungs, we iterate through each axial slice and find a minimum path [9] which separates the image slice into left and right halves while respecting the segmented object. A sample set of axial minimum paths is shown in Fig. 3(a). We then fit a smooth surface [10] to the set of axial minimizing paths which separates the whole lung segmented object into left and right components (Fig. 3(b)).

Note that the images acquired from asthmatics used in this study were diagnosed as mild to moderate. As such, the poorly ventilated regions were sufficiently minimal to allow for the described segmentation approach to work for all images in the study. We are currently pursuing a combined template strategy for more severe cases encountered in future studies.

2.3 Generation of Regional Masks

Not only is the calculation of image features over the whole lung essential for adequate lung characterization but feature generation over anatomically-based regional masks also provides important regional information unavailable to pulmonary function testing. Investigation into the distribution of airway disease in CT, such as emphysema, from the inner parenchymal core to the rind has demonstrated that the distribution of disease is generally more extensive in the core than in the rind region of the lung. Thus, corresponding inner core/outer rind regional masks are generated using binary morphological operations. We also further subdivide the left/right lungs into lower and upper portions for additional regional measurements. A total of 13 regional masks were created from the whole lung segmentations described in the previous section. The axial and coronal views of these regional masks are shown in Fig. 4.

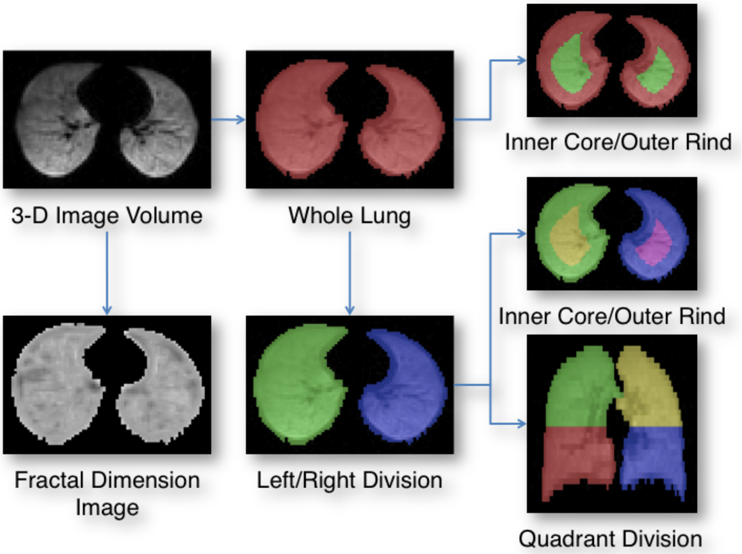


Fig. 4. Flowchart illustrating the generation of both feature images (e.g. stochastic fractal dimension image) and regional masks for localized analysis.

2.4 Image Features

For each of the 13 regional masks, three categories of features were calculated. These feature categories are:

- first order statistics derived from the intensity histogram of the original helium-3 image,
- first order statistics derived from the stochastic fractal dimension image [11], and
- second order statistics, or texture-based measurements, using the cooccurrence [12] and run-length [13] matrices of the original helium-3 image.

Helium-3 Intensity Histogram First Order Statistical Features. Within each of the regional masks, the grey-level intensity histogram is generated. This permits the calculation of the standard first order statistics, i.e. mean, variance, skewness, kurtosis, entropy, 5^{th}% intensity value, mean of 5^{th}%, and 95^{th}% intensity value, mean of 95^{th}%. From the histogram and image spacing we also calculate the volume of each lung.

Stochastic Fractal Dimension Image Features. Viewed as an intensity surface, Mandelbrot's fractal theory provides an informative framework for characterizing images [11, 14]. It has been shown that first order statistics derived from the stochastic fractal dimension (SFD) image facilitate discriminative feature analysis in CT images [1]. A SFD image is produced by iterating through the original image where, at each voxel, the corresponding fractal value requires inspection of each pair of voxels in the surrounding neighborhood. The average absolute intensity value difference is plotted against the voxel pair wise distances on a log-log scale. The voxel value is equal to 3 minus the slope of the line calculated using linear regression. First-order statistics from the SFD image in the various regions are calculated and used as image features.

Cooccurrence and Run-Length Matrix Features. Second order statistics have demonstrated utility in image texture classification. These include measurements derived from the grey-level cooccurrence matrix (CM) [12] and measurements derived from the run length matrix (RLM) [13]. As mentioned in the introduction, previous methodologies for evaluation of helium-3 MR images have included radiological assessment of "heterogeneity" for which these texture features are a surrogate. The specific set of CM features are energy, entropy, correlation, inverse difference moment, inertia, cluster shade, cluster prominence, and Haralick's correlation. The specific set of RLM features are short run emphasis, long run emphasis, grey level non-uniformity, run-length non-uniformity, run percentage, low grey level run emphasis, high grey level run emphasis, short run low grey level emphasis, short run high grey level emphasis, long run low grey level run emphasis, and long run high grey level run emphasis. This ordering of the CM and RLM features are used as subscripts in subsequent sections.

Otsu Thresholding of Parenchyma. Also utilized in previous research of hyperpolarized helium-3 MRI was radiological assessment of poor ventilation regional quantitation. Since the ventilation images do not reflect absolute regional ventilation we attempted to mimic this measurement in a simple fashion by applying an optimal thresholding to the grey-level histogram of the original image [15] to quantify regions of poor ventilation. We then divide the total volume of the ventilation defect regions by the total volume of the lungs to calculate a ventilation ratio.

2.5 Feature Characterization using Mutual Information

A straightforward approach to finding a subset of features most characterizing of the observed data is finding those features which correlate strongest with the target classification, i.e. the goal of "maximal relevance." The mutual information between two random variables x and y, defined as

$$I(x;y) = \int \int p(x,y) \log \frac{p(x,y)}{p(x)p(y)} dxdy, \tag{1}$$

is also used to define the dependency of variables. Given a set of N samples each characterized by M features $X = \{x_i, i = 1, \ldots, M\}$ where each sample is assumed to correspond to one of a finite number of classes described by the classification vector c, the maximally relevant feature, $x_i \in S \subset X$, maximizes the quantity $D = \frac{1}{|S|} \sum_{x_i \in S} I(x_i; c)$. In considering multiple features, it is desirable that the selected features be minimally redundant where redundancy is defined as $R = \frac{1}{|S|^2} \sum_{x_i, x_j \in S} I(x_i; x_j)$. The minimal-redundancy-maximal-relevance (mRMR) framework combines these two desiderata for feature subset selection by incrementally maximizing $\Phi = D - R$ [16].

3 Results

Hyperpolarized helium-3 images were acquired from 55 subjects (47 asthmatic and 8 healthy) before and after respiratory provocation (exercise or methacholine-induced). For each of the 110 images, a total of 533 features were calculated over all the anatomic regions. Each subject also underwent pulmonary function testing which generated an additional 27 clinical features, e.g. forced vital capacity (FVC), forced expiratory volume in 1 second (FEV_1), and peak expiratory flow (PEF). For each subject we combined the pre and post respiratory challenge images in addition to their difference values for a total of 1599 features. This was combined with the 27 PFT values for a total of 1626 features per subject.

Two rankings were generated from the mRMR algorithm using mutual information difference. For each entry, we show the rank, the calculated feature, whether the acquisition was pre or post respiratory provocation (or the difference of the two), the region, and the score. Table 1 ranks the relevancy of each feature considered individually whereas the ranking in Table 2 accounts for mutual redundancy between features. The top 10 individually considered features are given in Table 1. Since these features are strictly image-based, we also situate the top 3 PFT values within the rankings. We do a similar placement for Table 2.

Rank	Feature	Pre/Post/Diff	Region	Score
1	ventilation ratio	pre respiratory challenge	left lung	0.357
2	ventilation ratio	post respiratory challenge	left lung	0.329
3	RLM_5	post respiratory challenge	lower left lung	0.329
4	RLM_{10}	difference	whole lung	0.267
5	RLM_{10}	difference	outer rind, right lung	0.259
6	CM_2	post respiratory challenge	outer rind, left lung	0.198
7	CM_2	post respiratory challenge	lower left lung	0.193
8	RLM_5	post respiratory challenge	left lung	0.182
9	CM_6	post respiratory challenge	left lung	0.180
10	CM_4	post respiratory challenge	left lung	0.169
⋮	⋮	⋮	⋮	⋮
169†	% Predicted FEV_1	—	—	0.082
170†	FEV_3 Final	—	—	0.082
⋮	⋮	⋮	⋮	⋮
182†	% FEV_1	—	—	0.080

Table 1. Maximal relevance ranking (each feature is considered individually) of PFT and imaging features using the mRMR feature classification algorithm. Those rankings marked by a '†' denote a PFT value. The CM and RLM subscripts refer to the respective ordered measurement given in the text. Note the relative performance of the top image-based features compared with the top 3 PFT performers (rank 169, 170, and 182).

4 Discussion

There are several interesting observations to be made from the results. The first is rather obvious in that, considered as a global assessment, individual image features perform much better in characterizing clinical diagnosis compared with the PFT values. Another interesting observation is that despite the relatively poor individual classification performance of the PFT values, the better performers of this group have a minimal redundancy with respect to the image features. This is supported by the correlation values between the PFT results and each of the 1599 image features. For example the measurement % Predicted FEV_1 (rank 7 in Table 2) has correlation values with the image features in the range $[-0.4012, 0.4328]$ demonstrating practically no correlation with any of the image features. This demonstrates the orthogonal nature of the image-based information relative to the information provided by the PFT values.

Rank	**Feature**	**Pre/Post/Diff**	**Region**
1	ventilation ratio	pre respiratory challenge	left lung
2	RLM_5	post respiratory challenge	outer rind, left lung
3	ventilation ratio	post respiratory challenge	left lung
4	RLM_{10}	difference	right lung
5	CM_2	post respiratory challenge	lower left lung
6	RLM_8	post respiratory challenge	lower left lung
7†	% Predicted FEV_1	—	—
8	RLM_5	post respiratory challenge	lower left lung
9	CM_6	post respiratory challenge	outer rind, left lung
10	CM_2	post respiratory challenge	outer rind, whole lung
⋮	⋮	⋮	⋮
34†	FEV_1	—	—

Table 2. Maximal relevance, minimal redundancy ranking of PFT and imaging features using the mRMR feature classification algorithm. Those rankings marked by a '†' denote a PFT value. The CM and RLM subscripts refer to the respective ordered measurement given in the text.

It is also demonstrated that the ventilation ratio was the top classification feature, boding well for previous radiological assessments which attempted a similar calculation in quantifying regions of ventilation defects [4]. Also consistent with clinical understanding is the prominence of "post respiratory challenge" features amongst the top of the rankings since such provocation tends to exacerbate the asthmatic condition which is presumably reflected in the post images.

References

1. Xu, Y., van Beek, E.J.R., Hwanjo, Y., Guo, J., McLennan, G., Hoffman, E.A.: Computer-aided classification of interstitial lung diseases via MDCT: 3D adaptive multiple feature method (3d amfm). Acad Radiol **13**(8) (Aug 2006) 969–978
2. Sorenson, L., Shaker, S.B., de Bruijne, M.: Texture based emphysema quantification in lung ct. In: First International Workshop on Pulmonary Image Processing. (2008) 5–14
3. de Lange, E.E., Mugler, J.P., Brookeman, J.R., Knight-Scott, J., Truwit, J.D., Teates, C.D., Daniel, T.M., Bogorad, P.L., Cates, G.D.: Lung air spaces: Mr imaging evaluation with hyperpolarized 3he gas. Radiology **210**(3) (Mar 1999) 851–857
4. de Lange, E.E., Altes, T.A., Patrie, J.T., Parmar, J., Brookeman, J.R., Mugler, J.P., Platts-Mills, T.A.E.: The variability of regional airflow obstruction within the lungs of patients with asthma: assessment with hyperpolarized helium-3 magnetic resonance imaging. J Allergy Clin Immunol **119**(5) (May 2007) 1072–1078
5. Altes, T.A., Powers, P.L., Knight-Scott, J., Rakes, G., Platts-Mills, T.A., de Lange, E.E., Alford, B.A., Mugler, J.P., Brookeman, J.R.: Hyperpolarized 3he mr lung ventilation imaging in asthmatics: preliminary findings. J Magn Reson Imaging **13**(3) (Mar 2001) 378–384

6. Ibanez, L., Schroeder, W., Ng, L., Cates, J.: The ITK Software Guide. 2 edn. Kitware, Inc., Albany, NY (November 2005)
7. Sled, J.G., Zijdenbos, A.P., Evans, A.C.: A nonparametric method for automatic correction of intensity nonuniformity in mri data. IEEE Trans Med Imaging **17**(1) (Feb 1998) 87–97
8. Tustison, N.J., Gee, J.C.: N4ITK: Nick's N3 ITK implementation for MRI bias field correction. Insight Journal (2009)
9. Barrett, W.A., Mortensen, E.N.: Interactive live-wire boundary extraction. Medical Image Analysis **1**(4) (September 1997) 331–341
10. Tustison, N.J., Gee, J.C.: Generalized n-d C^k B-spline scattered data approximation with confidence values. In: Proc. Third International Workshop Medical Imaging and Augmented Reality. (2006) 76–83
11. Chen, C.C., DaPonte, J., Fox, M.: Fractal feature analysis and classification in medical imaging. IEEE Transactions on Medical Imaging **8**(2) (1989) 133–142
12. Haralick, R.M., Shanmugam, K., Dinstein, I.: Textural features for image classification. IEEE Transactions on Systems, Man, and Cybernetics **3**(6) (1973) 610–621
13. Dasarathy, B.R., Holder, E.B.: Image characterizations based on joint gray-level run-length distributions. Pattern Recognition Letters **12** (1991) 497–502
14. Tustison, N.J., Gee, J.C.: Stochastic fractal dimension image. Insight Journal (2009) published online
15. Otsu, N.: A thresholding selection method from gray-scale histogram. IEEE Transactions on System, Man, and Cybernetics **9** (1979) 62–66
16. Peng, H., Long, F., Ding, C.: Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy. IEEE Trans. Pattern Anal. Mach. Intell. **27**(8) (2005) 1226–1238