10-22-2019

Chronic Obstructive Pulmonary Disease: Thoracic CT Texture Analysis and Machine Learning to Predict Pulmonary Ventilation

Andrew Westcott

Dante P I Capaldi

David G McCormack

Aaron D Ward

Aaron Fenster

See next page for additional authors

Radiology

Chronic Obstructive Pulmonary Disease: Thoracic CT Texture Analysis and Machine Learning to Predict Pulmonary Ventilation

Andrew Westcott, MSc • Dante P. I. Capaldi, PhD • David G. McCormack, MD, FRCPC • Aaron D. Ward, PhD • Aaron Fenster, PhD, FCCPM • Grace Parraga, PhD

From the Robarts Research Institute, London, Canada (A.W., A.F., G.P.); Department of Medical Biophysics (A.W., A.D.W., A.F., G.P.), Division of Respirology, Department of Medicine (D.G.M., G.P.), and Department of Oncology (A.D.W.), Western University, 1151 Richmond St N, London, ON, Canada N6A 5B7; and Department of Radiation Oncology, Stanford University School of Medicine, Stanford, Calif (D.P.I.C.); Received February 26, 2019; revision requested May 17; final revision received August 6; accepted August 29. **Address correspondence to** G.P. (e-mail: *gparraga@robarts.ca*).

A.F. and G.P. supported by the Canadian Institutes of Health Research and the Natural Sciences and Engineering Research Council of Canada. A.F. supported by the Ontario Institute for Cancer Research. G.P. supported by the Canada Research Chairs program.

Conflicts of interest are listed at the end of this article.

See also the editorial by Fain in this issue.

Radiology 2019; 293:676–684 • https://doi.org/10.1148/radiol.2019190450 • Content codes: CH CT IN

Background: Fixed airflow limitation and ventilation heterogeneity are common in chronic obstructive pulmonary disease (COPD). Conventional noncontrast CT provides airway and parenchymal measurements but cannot be used to directly determine lung function.

Purpose: To develop, train, and test a CT texture analysis and machine-learning algorithm to predict lung ventilation heterogeneity in participants with COPD.

Materials and Methods: In this prospective study (*ClinicalTrials.gov*: NCT02723474; conducted from January 2010 to February 2017), participants were randomized to optimization (n = 1), training (n = 67), and testing (n = 27) data sets. Hyperpolarized (HP) helium 3 (^{3}He) MRI ventilation maps were co-registered with thoracic CT to provide ground truth labels, and 87 quantitative imaging features were extracted and normalized to lung averages to generate 174 features. The volume-of-interest dimension and the training data sampling method were optimized to maximize the area under the receiver operating characteristic curve (AUC). Forward feature selection was performed to reduce the number of features; logistic regression, linear support vector machine, and quadratic support vector machine classifiers were trained through fivefold cross validation. The highest-performing classification model was applied to the test data set. Pearson coefficients were used to determine the relationships between the model, MRI, and pulmonary function measurements.

Results: The quadratic support vector machine performed best in training and was applied to the test data set. Model-predicted ventilation maps had an accuracy of 88% (95% confidence interval [CI]: 88%, 88%) and an AUC of 0.82 (95% CI: 0.82, 0.83) when the HP ^{3}He MRI ventilation maps were used as the reference standard. Model-predicted ventilation defect percentage (VDP) was correlated with VDP at HP ^{3}He MRI (r = 0.90, $P < .001$). Both model-predicted and HP ^{3}He MRI VDP were correlated with forced expiratory volume in 1 second (FEV_1) (model: r = -0.65, $P < .001$; MRI: r = -0.70, $P < .001$), ratio of FEV_1 to forced vital capacity (model: r = -0.73, $P < .001$; MRI: r = -0.75, $P < .001$), diffusing capacity (model: r = -0.69, $P < .001$; MRI: r = -0.65, $P < .001$), and quality-of-life score (model: r = 0.59, P = .001; MRI: r = 0.65, $P < .001$).

Conclusion: Model-predicted ventilation maps generated by using CT textures and machine learning were correlated with MRI ventilation maps (r = 0.90, $P < .001$).

Online supplemental material is available for this article.

In chronic obstructive pulmonary disease (COPD), structural remodeling of the airways, airway inflammation or obliteration, and parenchyma destruction commonly result from chronic inhalation of combustible materials, including tobacco cigarettes and biomass fuels (1). Chest CT is used to visualize and quantify the many pulmonary structural abnormalities found in COPD. These measurements have been exploited in large cohort studies, including COPDGene (2), ECLIPSE (3), SPIROMICS (4), and CANCold (5), which have resulted in tens of thousands of thoracic CT images acquired in study participants with COPD (6). While all of these studies have focused on anatomic measurements, complementary functional information may also be gleaned by using inhaled xenon gas and dual-energy CT or multivolume CT acquisition through the breathing cycle in combination with registration and analysis techniques (7,8).

Hyperpolarized (HP) helium 3 (^{3}He) and xenon 129 (^{129}Xe) MRI pulmonary measurements also provide high spatial and temporal resolution of lung ventilation heterogeneity and microstructural information in COPD (9,10). HP gas MRI measurements in COPD are reproducible over short periods of time (11) and are sensitive to therapy (12) and to the lung changes that accompany exacerbations (13). In particular, HP ^{3}He and HP ^{129}Xe

Abbreviations

AUC = area under the receiver operating characteristic curve, CI = confidence interval, COPD = chronic obstructive pulmonary disease, CVM = cluster volume matrix, GLCM = gray-level co-occurence matrix, HP = hyperpolarized, SVM = support vector machine, VDP = ventilation defect percentage, VOI = volume of interest

Summary

In participants with chronic obstructive pulmonary disease, machine learning and texture analysis of chest CT were used to generate pulmonary ventilation maps that correlated with MRI ventilation maps and pulmonary function and quality-of-life measurements.

Key Results

- With use of machine learning of thoracic CT texture features to predict lung ventilation heterogeneity, overall accuracy was 88%, with an area under the receiver operating characteristic curve of 0.82.
- Ventilation maps based on conventional CT data were strongly correlated with MRI ventilation defect percentage (VDP) (r = 0.90, $P < .001$).
- Both model and MRI VDP were correlated with pulmonary function (for forced expiratory volume, model r = -0.65, $P < .001$; MRI r = -0.70, $P < .001$) and quality-of-life measurements (model r = 0.65, P = .001; MRI r = 0.59, $P < .001$).

MRI ventilation heterogeneity, quantified as MRI ventilation defect percentage (VDP) (14), is predictive of COPD exacerbations (15) and longitudinal changes in quality of life and exercise capacity (16). Despite these unique advantages, HP gas MRI has been limited to specialized research centers and has not been used in multicenter cohort COPD studies and clinical trials, largely because of the cost of these HP noble gases and the specialized equipment needed (specialty tuned surface coils and a noble gas hyperpolarizer) for MRI ventilation imaging. For these reasons, the unique functional information provided by HP gas MRI has not been translated to the clinic. With the recent success of texture analysis and machine learning in medical imaging (17,18), we postulated that it would be possible to identify sufficient features in CT images to generate lung ventilation heterogeneity maps, which would make this important lung functional information more widely available.

The TINCan cohort study (19) prospectively acquired volume-matched CT and MRI scans in a relatively large group of study participants with COPD and provided a unique opportunity to train and test a machine-learning approach to generate pulmonary ventilation maps based on single-volume, noncontrast CT data. We hypothesized that ventilation maps could be generated based on texture features and machine learning of CT images, and that such maps would correlate spatially with HP ^{3}He MRI VDP acquired experimentally in participants with COPD. Such pulmonary ventilation predictions based on nearly universally available conventional CT may increase clinical access to valuable functional lung information beyond sites with HP gas MRI. Therefore, the purpose of this study was to develop, train, and test a pulmonary CT texture analysis and machine learning pipeline to predict HP ^{3}He MRI ventilation heterogeneity maps acquired in the same participants with COPD.

Materials and Methods

Participants provided written informed consent to this prospective, longitudinal cohort study (20), which was approved by a local research ethics board in compliance with the Health Insurance Portability and Accountability Act (*ClinicalTrials.gov*: NCT02723474; Institutional Review Board #00000940) (19). The authors had control of the data and information submitted for publication. Participants evaluated in this study have been reported previously (16,19); however, the analysis undertaken in the current study was unique and has never been reported before in this group of COPD participants or in any other cohort study.

Study Participants

Participants with a clinical diagnosis of COPD between the ages of 40 and 85 years were recruited from a tertiary-care academic center in London, Ontario, Canada between 2010 and 2017 as a convenience sample. The Consolidated Standards of Reporting Trials diagram is provided in Figure 1. Participants were excluded if they did not have COPD or if the CT acquisition parameters were prospectively altered to include a modified inspiration-expiration protocol or scheme. One participant data set was used to perform texture parameter optimization, and the remaining participant data sets were randomized to a training set (for tuning model hyperparameters) and a testing set.

Study Design and Pulmonary Function Tests

Spirometry measurements were acquired according to American Thoracic Society guidelines (21) by using a whole-body plethysmography system (MedGraphics, St Paul, Minn), and results were corrected for age and sex (percentage predicted). Body plethysmography was performed for the measurement of lung volumes, and diffusing capacity of lung for carbon monoxide was measured by using the attached gas analyzer. The St George's Respiratory Questionnaire (22) was used to measure participant quality of life.

MRI Examination

Conventional proton (hydrogen 1 [^{1}H]) and HP ^{3}He MRI were performed with a whole-body 3.0-T MRI system (MR750 Discovery; GE Healthcare, Waukesha, Wis) with broadband imaging capabilities as previously described (9). Hydrogen 1 MRI was performed by using a fast spoiled gradient-recalled-echo sequence, with acquisition parameters as previously described (9). HP ^{3}He MRI involved the use of a whole-body gradient set with maximum gradient amplitude of 50 mT/m and a single-channel, rigid elliptical transmit-receive chest coil (RAPID Biomedical, Wuerzburg, Germany). The basis frequency of the coil was 97.3 MHz, and the excitation power was 3 kW using an AMT 3T90 RF power amplifier (GE Healthcare). The ^{3}He gas was polarized to 30%–40% polarization using a spin-exchange optical polarizer (Polarean, Durham, NC).

CT Examination

Chest CT was performed with a 64-section Lightspeed VCT scanner (GE Healthcare) (64 × 0.625 mm, 120 kVp, 100 mA [effective], tube rotation time = 500 msec, and pitch = 1.0). CT images were reconstructed by using a section thickness of

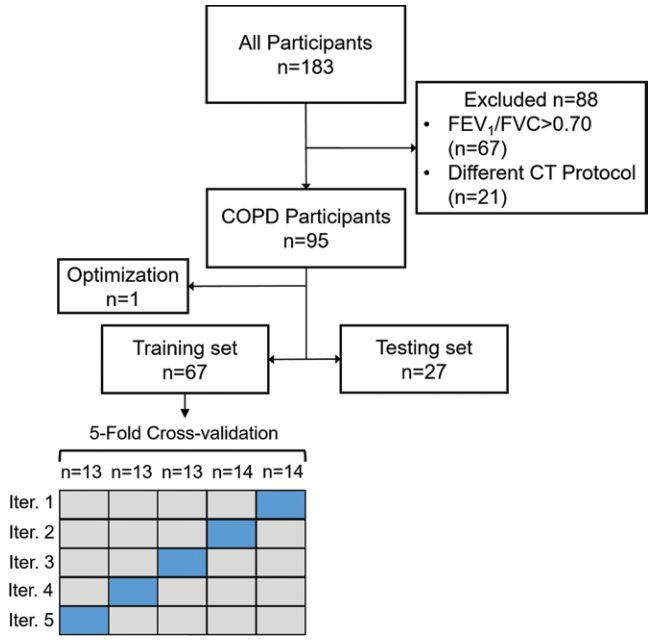

Figure 1: Consolidated Standards of Reporting Trials (CONSORT) diagram shows the number of participants who completed a visit and the number of participants excluded in this study because they did not have chronic obstructive pulmonary disease (COPD) or because their CT data were acquired with different voxel spacing. For the training set, participants were randomly divided into five groups. Training was performed during five iterations, whereby for each iteration the model was trained on four groups (gray rectangles) and validated on one group (blue rectangles). FEV_1 = forced expiratory volume in 1 second, FVC = forced vital capacity, Iter. = iteration.

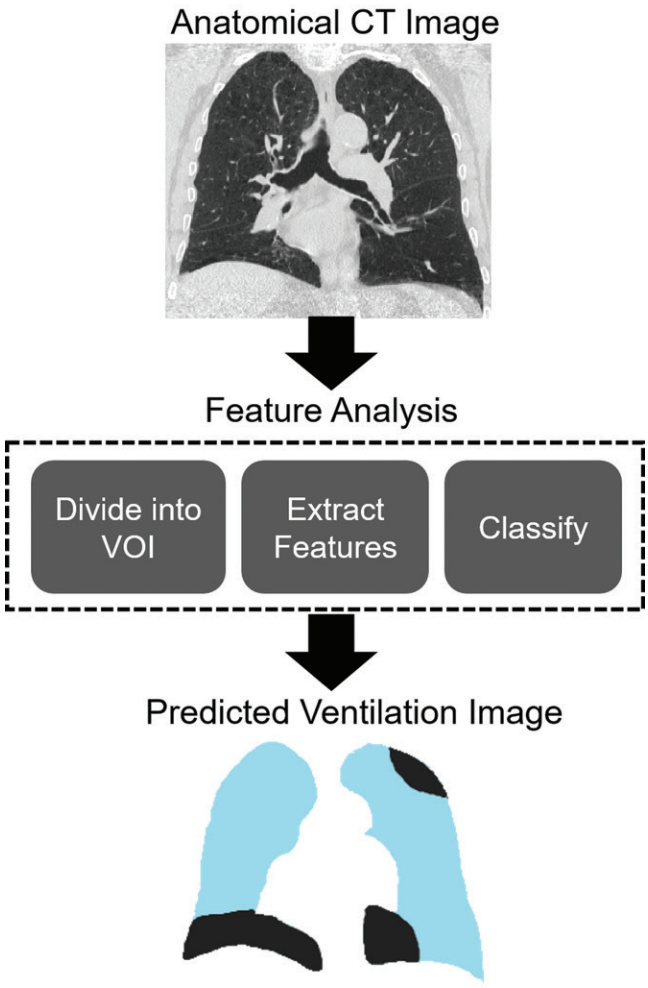

Figure 2: Schematic for image analysis shows texture analysis image processing steps for generating predicted ventilation maps from thoracic CT for an 82-year-old man with chronic obstructive pulmonary disease (forced expiratory volume in 1 second [FEV_1] = 60% predicted value; FEV_1/forced vital capacity = 45%; diffusing capacity of lung for carbon monoxide = 33% predicted value). In the predicted ventilation image, aqua = ventilated regions and dark gray = ventilation defects. VOI = volume of interest.

1.25 mm (0.5 pitch) with a standard convolution kernel. The total effective dose to the participant was 1.8 mSv according to manufacturer settings and the Imaging Performance Assessment of CT patient dosimetry calculator based on software from the Health Protection Agency of the United Kingdom (NRPB-SR250). The voxel dimensions in the coronal plane were 3.1 × 3.1 × 15 mm^3 for MRI and 1.25 × 0.7 × 0.7 mm^3 for CT. Both MRI and CT were performed during a static breath hold at functional residual capacity plus 1 L.

Image Coregistration

Figure 2 shows the image processing pipeline we developed to generate ventilation maps based on thoracic CT data. Figure 3 shows the MRI-CT registration and the volume-of-interest (VOI) extraction approach. MRI ^1H and ^3He images were registered by using landmark registration, and k-means clustering was used to generate ventilation cluster maps, as previously described (14), with the lowest cluster representing ventilation defects. MRI data were resampled by using nearest neighbors to 1.25 × 0.7-mm voxels to match the original CT coronal plane dimensions and were cropped to match the CT field of view. CT images were segmented by using Pulmonary Workstation 2.0 (VIDA Diagnostics, Coralville, Iowa) and then concatenated to 15-mm-thick sections in the coronal plane to match the MRI dimensions. The segmented ^1H MRI thoracic cavity masks were registered to the CT data by using a deformable registration method (modality independent neighborhood descriptor, or MIND, registration) (23). The resultant registration transformation was applied to the MRI ventilation cluster maps. Coregistration was evaluated by using the Dice similarity coefficient.

CT and MRI VOIs were defined by applying a three-dimensional grid, with dimensions of $L \times L \times 15$ mm^3, to the segmented CT and MRI ventilation map, where L was the VOI size in the coronal plane, which was optimized within the training set. The MRI ventilation map was labeled as background, ventilated, or nonventilated, with the label of each VOI being the mode, or most common, value. The grid was then shifted $L/2$ mm horizontally and $L/2$ mm vertically, to generate additional training samples, such that each voxel belonged to three separate VOIs. This technique was further used when predicting the final label in the test set, as the mean score of three overlapping $L \times L \times 15$ VOIs was used, which then defined unique predictions for $L/2 \times L/2 \times 15$ mm^3 regions.

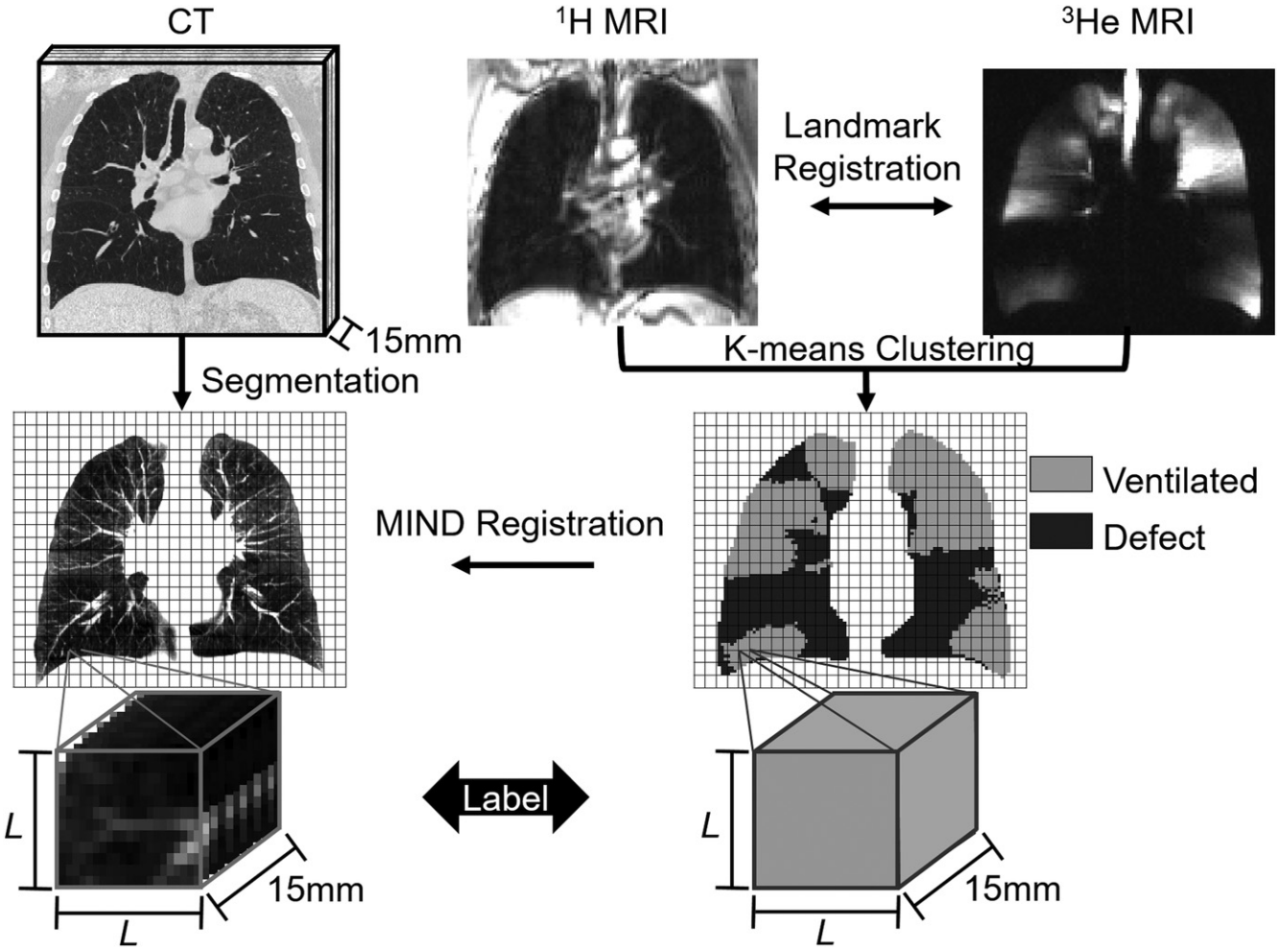

Figure 3: MRI-to-CT registration analysis. A three-dimensional grid was used to define the CT volume of interest (VOI) and the corresponding MRI-based ventilation label, where *L* is the length of the VOI in the coronal plane, which is varied to determine the optimal VOI dimensions. Thoracic images are shown for a 75-year-old man with chronic obstructive pulmonary disease (forced expiratory volume in 1 second [FEV_1] = 28% predicted value; FEV_1/forced vital capacity = 29%; diffusing capacity of lung for carbon monoxide = 17% predicted value). MIND = modality independent neighborhood descriptor.

Thoracic CT Feature Extraction

First- and second-order features were extracted by using a custom-built texture analysis software in MATLAB (MATLAB R2018a; MathWorks, Natick, Mass) available from the authors online *(http://www.imaging.robarts.ca/parraga/our_code.html)*. A parameter search was performed to determine the texture parameters (gray-level co-occurence matrix [GLCM] bin width, cluster volume matrix [CVM] bin width, and CVM bin range) by using one participant data set, which was then removed for the remainder of the analysis. The results from this parameter search are provided in Figure E1 (online). Exemplar feature maps using these optimized parameters are shown in Figure 4. There were 87 global features calculated per VOI. To provide the model more context in terms of COPD severity, each VOI feature was also divided by the average value of the feature within the same participant, and this generated an additional 87 features (ie, ratio features), for a total of 174 features per VOI.

First-order features that were generated included mean CT attenuation, standard deviation of attenuation, skewness, kurtosis, 2nd moment, 3rd moment, 95th percentile attenuation, 15th percentile attenuation, relative area (RA) of the lung less than –950 HU (RA_{950}), RA less than –910 HU (RA_{910}), and RA less than –856 HU (RA_{856}).

The GLCM (24) was populated by binning voxels based on Hounsfield units into 45 bins from 0 to –1000 HU, where 45 bins was the result from the parameter search in the single participant's data set removed participant data set. As previously described (24), features were calculated as shown in the right side of Table E1 (online) for the 13 unique three-dimensional directions.

Although run-length matrix is commonly used as a method to extract features from two-dimensional images (25), this becomes computationally intensive and does not account for regions that are fully but not linearly connected, which can be accounted for by using low-attenuating clusters (26). Therefore, here we used a combination of the run-length matrix with CT cluster analysis by creating a new texture parameter, the CVM. This is a three-dimensional analog of the run-length matrix, where $p(i,j)$ is the number of clusters, i is the gray level, and j is the three-dimensional cluster size of the same gray level. The features calculated from the CVM are the same as those calculated from a run-length matrix, as shown on the left side of Table

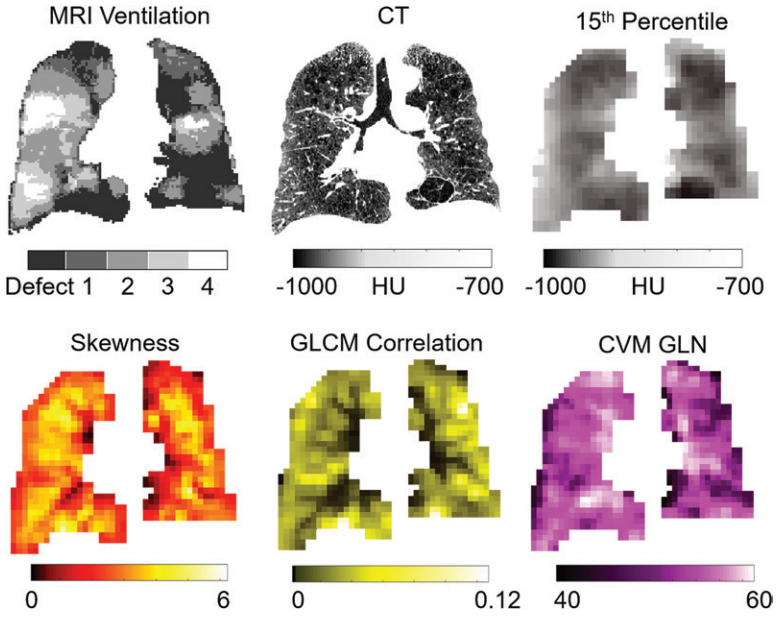

Figure 4: Hyperpolarized MRI ventilation, thoracic CT image, and feature maps in an 86-year-old man with chronic obstructive pulmonary disease (forced expiratory volume in 1 second [FEV_1] = 38% predicted value; FEV_1/forced vital capacity = 35%; diffusing capacity of lung for carbon monoxide = 36% predicted value). Feature maps were calculated by using volume-of-interest dimensions of 15 × 15 × 15 mm^3. CVM = cluster volume matrix, 15th percentile = the 15th percentile of the CT attenuation histogram, GLCM = gray level co-occurrence matrix, GLN = gray-level nonuniformity.

Table 1: Participant Demographic Data and Pulmonary Function and Imaging Measurements

Parameter	All Participants (n = 94)	Training Set (n = 67)	Test Set (n = 27)	P Value
Age (y)	70 ± 9	70 ± 8	69 ± 10	.6
Percentage of women	33	31	48	.2
No. of pack-years smoked	50 ± 30	49 ± 29	50 ± 27	.9
FEV_1 (percentage predicted)	63 ± 25	63 ± 25	61 ± 24	.6
FEV_1/FVC (percentage predicted)	51 ± 13	51 ± 13	51 ± 13	>.99
DLco (percentage predicted)	56 ± 23*	54 ± 21)†	57 ± 22	.7
SGRQ score	40 ± 18‡	41 ± 18)§	37 ± 20‖	.5
RA_{950} (%)	10 ± 10	10 ± 10	10 ± 10	.7
VDP	12 ± 12	12 ± 11	11 ± 12	.9
DSC (%)	95 ± 1	95 ± 1	95 ± 1	.6

Note.—Unless otherwise specified, data are means ± standard deviations. Pack-years = no. of cigarette packs smoked per day times years of smoking. DLco = diffusing capacity of lung for carbon monoxide, DSC = Dice similarity coefficient for CT-MRI coregistration, FVC = forced vital capacity, FEV_1 = forced expiratory volume in 1 second, RA_{950} = relative area of CT histogram less than –950 HU, SGRQ = St George's Respiratory Questionnaire, VDP = ventilation defect percentage.

* Available for 93 participants.

† Available for 66 participants.

‡ Available for 90 participants.

§ Available for 64 participants.

‖ Available for 26 participants.

E1 (online). The matrix was populated by binning voxels into 18 bins between –300 and –1000 HU, where all voxels with values greater than –300 were collapsed into a single bin and all voxels less than –1000 HU were included in the lowest density bin.

Feature Selection

To avoid model overfitting and maximize model generalizability, a forward feature selection scheme was developed in MATLAB. As shown in Figure 1, training data were divided into five different groups, and fivefold cross validation was performed by using logistic regression, where features were iteratively added based on the feature that led to the greatest improvement in the validation area under the receiver operating characteristic curve (AUC). As shown in Figure E2, *A* (online), this was performed for up to 20 features, as the AUC plateaued to a maximum when 20 features were used. Those features, which were included in the first 20 features selected for at least two of the fivefold cross-validation steps, were utilized in the final model approach provided by the authors online *(http://www.imaging.robarts.ca/parraga/our_code.html)*.

Training the Classification Model

To simplify the anatomic heterogeneity included in the model while maintaining the information that was representative of the entire lung, we included the center-most six sections (of approximately 15 total sections) that encompassed 58% ± 5 of the volume of the lung. In the training set, the VOI size and sampling scheme were optimized to maximize AUC. The VOI dimensions were 15 mm in the posterior-to-anterior direction and were varied from 15 × 15 to 30 × 30 mm^2 in the coronal plane. Because of a greater number of ventilated versus nonventilated VOIs, the method for sampling VOIs to create the training set was critical. To include a balance of participants with large ventilation defects and participants with a lower magnitude of VDP, a threshold was varied to define the minimum percentage of each participant's lung sampled (ie, for threshold = 20%, participants with VDP < 20%, ventilated volume sampled = 20% of the lung and nonventilated volume sampled = VDP). The sampling scheme that resulted in the highest AUC was used to train the final model.

Once all parameters and features were selected, fivefold cross-validation training was performed by using a logistic regression (with no interactions considered), a linear support vector machine (SVM), and a quadratic SVM. The data were standardized and hyperparameter optimization was performed by using MATLAB (Classification Learner app) for

Table 2: VOI and Training Sample Patterns

Parameter	Validation Accuracy (%)	Validation AUC	Validation Sensitivity (%)	Validation Specificity (%)
Coronal VOI size (mm)				
15 × 15	82 (81, 82)	0.77 (0.77, 0.78)	44 (42, 46)	89 (88, 89)
20 × 20	79 (78, 79)	0.78 (0.78, 0.79)	58 (56, 60)	82 (82, 83)
25 × 25	80 (79, 81)	0.80 (0.80, 0.81)	61 (58, 63)	83 (83, 84)
30 × 30	85 (84, 86)	0.82 (0.81, 0.83)	49 (47, 52)	91 (90, 91)
Sampling scheme minimum percentage of lung sampled				
0	69 (68, 70)	0.78 (0.77, 0.78)	73 (70, 76)	69 (68, 70)
10	79 (78, 80)	0.81 (0.80, 0.82)	66 (64, 68)	81 (80, 82)
20	83 (82, 83)	0.82 (0.81, 0.82)	58 (55, 61)	86 (86, 87)
30	85 (84, 86)	0.82 (0.81, 0.83)	49 (47, 52)	91 (90, 91)
40	86 (86, 87)	0.82 (0.81, 0.83)	38 (35, 41)	94 (93, 94)

Note.—Data in parentheses are 95% confidence intervals. To optimize coronal volume of interest (VOI) dimensions, the sampling scheme ensured that 30% or more of the ventilated lung was evaluated. To optimize the sampling scheme, for each participant evaluated in the training set, at least as many ventilated samples as nonventilated samples were evaluated. To ensure that disease severity (ventilation, ventilation defects) was represented to train the model, a minimum percentage of the ventilated lung was sampled. AUC = area under the receiver operating characteristic curve.

Table 3: Training Set Features and Individual AUCs for 67 Participants and 48 313 VOIs

Feature	AUC
15th percentile global	0.76 (0.76, 0.77)
Skewness global	0.51 (0.50, 0.51)
15th percentile ratio	0.71 (0.71, 0.72)
GLCM energy [-1,1,1] ratio	0.58 (0.58, 0.59)
GLCM homogeneity [0,1,0] ratio	0.57 (0.56, 0.58)
Skewness ratio	0.51 (0.50, 0.52)
CVM GLN ratio	0.62 (0.61, 0.63)
RA_{950} ratio	0.71 (0.70, 0.72)
CVM LGRE global	0.55 (0.54, 0.56)
RA_{950} global	0.76 (0.75, 0.77)
GLCM energy [1,1,-1] ratio	0.59 (0.58, 0.60)
GLCM energy [1,0,0] global	0.58 (0.58, 0.59)
GLCM contrast [1,0,1] global	0.59 (0.58, 0.60)
RA_{910} global	0.75 (0.75, 0.76)
CVM GLN global	0.66 (0.65, 0.67)
GLCM energy [1,1,0] ratio	0.58 (0.57, 0.59)
GLCM contrast [1,0,-1] ratio	0.59 (0.59, 0.60)
CVM SRLGE global	0.55 (0.54, 0.56)
CVM HGRE global	0.57 (0.56, 0.58)
GLCM correlation [1,0,-1] ratio	0.59 (0.59, 0.60)
Standard deviation ratio	0.54 (0.53, 0.54)

Note.—Data in parentheses are 95% confidence intervals. AUC = area under the receiver operating characteristic curve, CVM = cluster volume matrix, 15th percentile = 15th percentile of CT, GLCM = gray-level co-occurrence matrix, GLN = gray-level nonuniformity, global = value of volume of interest (VOI) feature, HGRE = high gray-level run emphasis, LGRE = low gray-level run emphasis, RA_{950} = relative area of CT histogram less than −950 HU, ratio = value of VOI feature divided by mean feature value for whole lung, SRLGE = short run low gray-level emphasis.

each model to find the optimal hyperparameters, as shown in Table E2 (online). The mean cross-validation AUCs were used to compare models, and the model with the greatest AUC was applied to the testing set. There was no patient overlap between the training and testing data sets, as well as no patient overlap between each of the five validation experiments. This ensured that the model was trained on a data set that was different from the testing data set.

Statistical Analysis

All statistical analyses were performed by using GraphPad Prism, V8.0 (GraphPad Software, La Jolla, Calif). To determine group differences, unpaired parametric *t* tests were used when the data were normally distributed, and nonparametric Mann-Whitney tests were used when the data were not normally distributed (tested by using the Shapiro-Wilk test for normality). Pearson correlation coefficients were used to determine the relationships between the model and the HP ^{3}He ventilation MRI and pulmonary function measurements. Accuracy, sensitivity, specificity, and AUC were calculated on a VOI level. Sensitivity, specificity, and AUC were the primary metrics of performance and were the metrics considered during training. Results were considered significant when the probability of a two-tailed type I error (α) was less than 5% ($P < .05$).

Results

Of the 183 participants enrolled in the study, 88 were excluded for not having COPD ($n = 67$) or for having different CT acquisition parameters ($n = 21$). The participants included in this study were 33% women, with a mean age of 70 years ± 9 and a mean smoking history of 50 years ± 30. Complete demographic data, pulmonary function test results, and imaging measurements are provided in Table 1 and show there were no differences between the training and testing data sets.

As shown in Table 2, the best-performing VOI size was 30 × 30 mm^2 based on the largest AUC. The optimal sampling threshold required a minimum of 30% ventilated lung, based on the AUC while maintaining the sensitivity once the AUC plateaued. Table 2 also shows that by varying the sampling pattern, there was a trade-off between sensitivity and specificity. Therefore, feature selection was performed by using fivefold cross validation with VOI dimensions of 30 × 30 × 15 mm^3 and a sampling minimum of 30%.

Table 3 shows that first-order density-based features had the largest individual AUC and both global and ratio values helped

Figure 5: Representative CT, MRI, and model outputs for four study participants in the testing set. Images in participants with a range of chronic obstructive pulmonary disease severities are shown, where qualitative spatial correlations between predicted ventilation and measured ventilation are provided. Participant 1 was an 83-year-old man (forced expiratory volume in 1 second [FEV_1] = 116% predicted value, FEV_1/forced vital capacity [FVC] = 67%, diffusing capacity of lung for carbon monoxide [DLco] = 107% predicted value), participant 2 was a 66-year-old woman (FEV_1 = 39% predicted value, FEV_1/FVC = 34%, DLco = 63% predicted value), participant 3 was a 75-year-old man (FEV_1 = 25% predicted value, FEV_1/FVC = 29%, DLco = 17% predicted value), and participant 4 was a 75-year-old man (FEV_1 = 30% predicted value, FEV_1/FVC = 30%, DLco = 39% predicted value).

optimize the model, where a total of 21 features were included in the final model. The correlation plot for all included features (Fig E2, *B* [online]) shows the strong correlation between GLCM features in different directions. During training, logistic regression, linear SVM, and quadratic SVM achieved accuracies and AUCs, respectively, of 85% (95% confidence interval [CI]: 84%, 86%) and 0.82 (95% CI: 0.81, 0.83), 86% (95% CI: 85%, 87%) and 0.81 (95% CI: 0.80,0.82), and 87% (95% CI: 86%, 88%) and 0.86 (95% CI: 0.85,0.87), respectively. Therefore, the quadratic SVM model was applied to the test set based on the increased AUC.

Figure 5 shows the predicted ventilation maps for four participants within the test set, along with the experimentally acquired CT and MRI ventilation images. A qualitative spatial agreement between the model ventilation prediction and the MRI ventilation was observed, and the magnitude of predicted ventilation defects corresponds to that observed in the MRI ventilation.

In Figure 6, test set evaluations are shown where the final model achieved an accuracy of 88% (95% CI: 88%, 88%) and an AUC of 0.82 (95% CI: 0.82, 0.83). Figure 6 also shows the strong relationship between model-predicted VDP and HP MRI VDP (r = 0.90, $P \leq .0001$) as well as their relationships with forced expiratory volume in 1 second (FEV_1) (model: r = -0.65, $P < .001$; MRI: r = -0.70, $P < .0001$), FEV_1/forced vital capacity (model: r = -0.73, $P < .0001$; MRI: r = -0.75, $P < .0001$), diffusing capacity of lung for carbon monoxide (model: r = -0.69, $P < .0001$; MRI: r = -0.65, $P < .001$) and St George's Respiratory Questionnaire score (model: r = 0.59, $P < .005$; MRI: r = 0.65, $P < .001$).

Discussion

We developed a CT analysis pipeline that combined texture feature analysis with machine learning to generate pulmonary ventilation heterogeneity maps for direct comparison with hyperpolarized (HP) helium 3 (^{3}He) MRI ventilation maps acquired in study participants. The algorithm was trained and tested in a cohort of 95 study participants with chronic obstructive pulmonary disease (COPD) in whom volume-matched MRI and CT were performed within 10 minutes of each other. We made the following observations: *(a)* The best-performing model on the training set was applied to the testing set, where it classified ventilated and nonventilated volume of interest with 88% accuracy and an area under the receiver operating characteristic curve (AUC) of 0.82; *(b)* there was a strong correlation between model-predicted ventilation defect percentage (VDP) and HP ^{3}He MRI VDP (r = 0.90, $P < .001$); and *(c)* both model-predicted and HP ^{3}He MRI VDP were correlated with clinically relevant measurements such as forced expiratory volume in 1 second/forced vital capacity

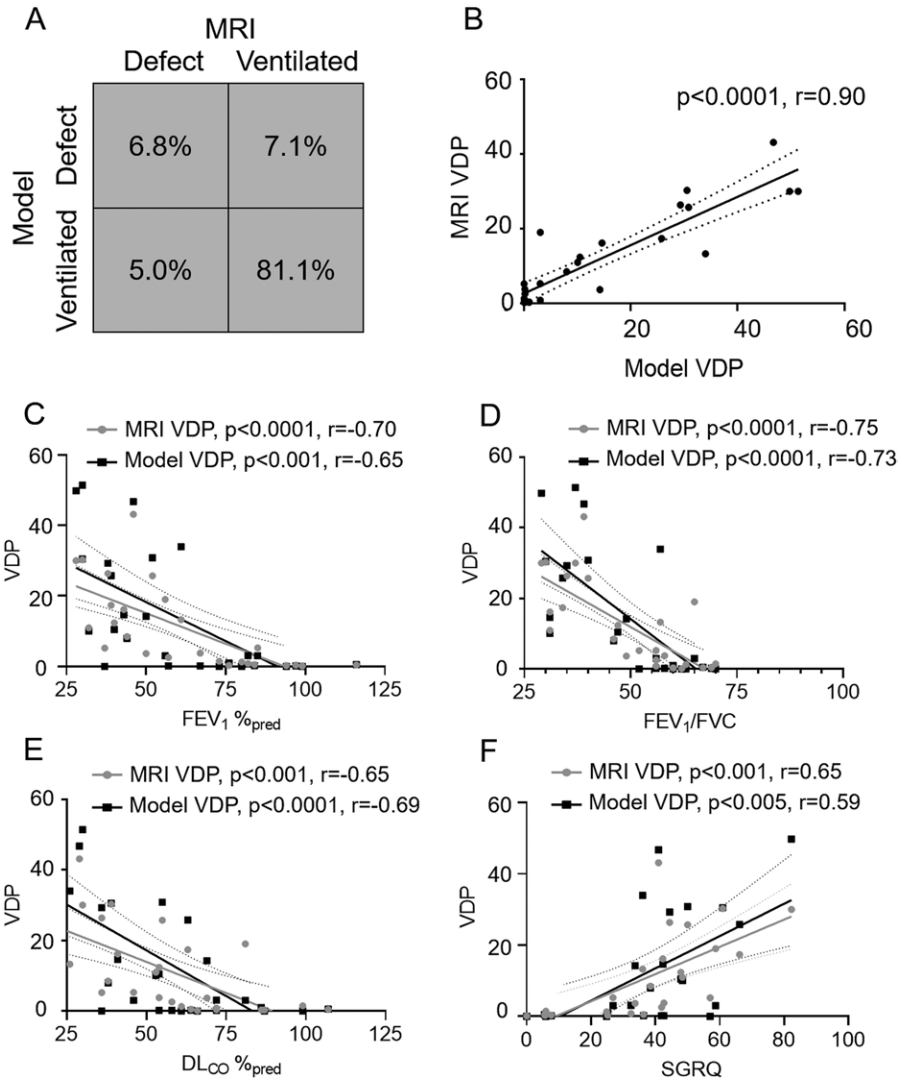

Figure 6: Testing set model outputs and relationships. *A*, Confusion matrix shows the model-predicted and the ground-truth hyperpolarized helium 3 MRI ventilation classification, where accuracy was 88% (95% confidence interval [CI]: 88%, 88%), the area under the receiver operating characteristic curve was 0.82 (95% CI: 0.82, 0.83), the sensitivity was 58% (95% CI: 56%, 59%), and the specificity was 92% (95% CI: 92%, 92%) in 32 457 volumes of interest. *B*, Graph shows predicted ventilation defect percentage (VDP) versus observed VDP ($r = 0.90$, $y = 0.65x + 3$). *C–F*, Graphs show clinical measurements versus model VDP and MRI VDP. DLco = diffusing capacity of the lung for carbon monoxide, FEV_1 = forced expiratory volume in 1 second, FVC = forced vital capacity, SGRQ = St George's Respiratory Questionnaire.

(model $r = -0.73$, $P < .0001$; MRI: $r = -0.75$, $P < .0001$), diffusing capacity of lung for carbon monoxide (model: $r = -0.69$, $P < .0001$; MRI: $r = -0.65$, $P < .001$), and St George's Respiratory Questionnaire score (model: $r = 0.59$, $P < .005$; MRI: $r = 0.65$, $P < .001$).

Because of the imbalance between ventilated and nonventilated volumes, we identified the parameters and final model on the basis of maximizing the AUC, such that both sensitivity and specificity of detecting ventilation heterogeneities were considered for algorithm performance. However, the sensitivity and specificity of the final model were not balanced, allowing for a higher specificity to minimize the false detection of ventilation defects. Attenuation-based first-order features generated the largest AUCs, which was consistent with the spatial overlap between emphysema and ventilation defects and in agreement with previous experimental results (15,27). The influence of both global and ratio features for predicting ventilation underscores the importance of considering disease severity in relation to the rest of an individual's lung and relative to all individual lungs.

The strong relationship between the model, pulmonary function test results, and MRI VDP also provides support that this model would predict a wide range of disease severity present within our study. Our results are important in the context of previous automated disease quantification methods developed by using texture analysis (28–31), which were trained by using unsupervised learning or with previously developed disease classification systems. In contrast, our predicted model provided a quantitative measure that was spatially dependent and trained by using HP ^{3}He MRI ventilation results as the ground truth. The bias to overestimate ventilation defects was an artifact of training the model on a data set with a more balanced ventilated-to-nonventilated volume ratio compared with the ratio present within the testing set.

We acknowledge numerous study limitations, including the fact that the TINCan cohort comprised a convenience sample with many patients with moderate to severe disease, so generalizability should be considered in this context. We also acknowledge that only the center 60% or so of the lung was included in the analysis to simplify CT-MRI coregistration. Nevertheless, the center six sections provided an appropriate representation of all five lung lobes, with a mean VDP difference of 1% ± 2 between the sections we used and whole-lung measurements. CT protocol and scanner image reconstruction may influence texture features (32,33), and hence, the generalizability of the trained model presented here to other scanners and CT acquisition protocols needs to be considered. Further, MRI and CT were both performed at the same lung volume (functional residual capacity plus 1 L), and the participants were coached to ensure that this was the case. Lung quantitative features are known to be influenced by lung volume, which has implications for the general application of this approach (34). With that in mind, it is expected that the inclusion of the normalized features (to whole-lung means) may help mitigate interscanner and intersite variability. While application of our approach to a multicenter data set will evaluate its utility, it is worth noting

that many of the density-based measures such as relative area of the lung less than −950 HU and 15th percentile attenuation are already successfully used in clinic as well as in multisite studies (35,36).

In conclusion, in participants with chronic obstructive pulmonary disease, machine learning and texture analysis of chest CT data was used to generate pulmonary ventilation maps that correlated with MRI ventilation maps and pulmonary function and quality-of-life measurements. This approach, if externally validated, will enable widespread generation of ventilation heterogeneity maps using nearly ubiquitous CT scanners, providing a way to generate ventilation maps beyond the specialist centers with hyperpolarized gas MRI.

Author contributions: Guarantors of integrity of entire study, D.P.I.C., G.P.; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; approval of final version of submitted manuscript, all authors; agrees to ensure any questions related to the work are appropriately resolved, all authors; literature research, D.P.I.C., G.P.; clinical studies, D.P.I.C., G.P.; experimental studies, A.W., D.P.I.C., D.G.M.; statistical analysis, A.W., D.P.I.C.; and manuscript editing, all authors

Disclosures of Conflicts of Interest: **A.W.** disclosed no relevant relationships. **D.P.I.C.** disclosed no relevant relationships. **D.G.M.** disclosed no relevant relationships. **A.D.W.** disclosed no relevant relationships. **A.F.** Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: is about to file a patent on a counterbalanced mechanical system with WORLDiscoveries; institution's commercialization office has licensed A.F.'s patents for 3D US and distributes money from royalties to investors, who include A.F.; owns pension plan that invests money in stocks and bonds. Other relationships: disclosed no relevant relationships. **G.P.** disclosed no relevant relationships.

References

1. Vogelmeier CF, Criner GJ, Martinez FJ, et al. Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Lung Disease 2017 Report. GOLD Executive Summary. Am J Respir Crit Care Med 2017;195(5):557–582.
2. Regan EA, Hokanson JE, Murphy JR, et al. Genetic epidemiology of COPD (COPDGene) study design. COPD 2010;7(1):32–43.
3. Vestbo J, Anderson W, Coxson HO, et al. Evaluation of COPD Longitudinally to Identify Predictive Surrogate End-points (ECLIPSE). Eur Respir J 2008;31(4):869–873.
4. Couper D, LaVange LM, Han M, et al. Design of the Subpopulations and Intermediate Outcomes in COPD Study (SPIROMICS). Thorax 2014;69(5):491–494.
5. Bourbeau J, Tan WC, Benedetti A, et al. Canadian Cohort Obstructive Lung Disease (CanCOLD): Fulfilling the need for longitudinal observational studies in COPD. COPD 2014;11(2):125–132.
6. González G, Ash SY, Vegas-Sánchez-Ferrero G, et al. Disease staging and prognosis in smokers using deep learning in chest computed tomography. Am J Respir Crit Care Med 2018;197(2):193–203.
7. Hoffman EA, Lynch DA, Barr RG, van Beek EJ, Parraga G; IWPFI Investigators. Pulmonary CT and MRI phenotypes that help explain chronic pulmonary obstruction disease pathophysiology and outcomes. J Magn Reson Imaging 2016;43(3):544–557.
8. Tahir BA, Hughes PJC, Robinson SD, et al. Spatial Comparison of CT-Based Surrogates of Lung Ventilation With Hyperpolarized Helium-3 and Xenon-129 Gas MRI in Patients Undergoing Radiation Therapy. Int J Radiat Oncol Biol Phys 2018;102(4):1276–1286.
9. Parraga G, Ouriadov A, Evans A, et al. Hyperpolarized 3He ventilation defects and apparent diffusion coefficients in chronic obstructive pulmonary disease: preliminary results at 3.0 Tesla. Invest Radiol 2007;42(6):384–391.
10. Kirby M, Svenningsen S, Kanhere N, et al. Pulmonary ventilation visualized using hyperpolarized helium-3 and xenon-129 magnetic resonance imaging: differences in COPD and relationship to emphysema. J Appl Physiol (1985) 2013;114(6):707–715.
11. Parraga G, Mathew L, Etemad-Rezai R, McCormack DG, Santyr GE. Hyperpolarized 3He magnetic resonance imaging of ventilation defects in healthy elderly volunteers: initial findings at 3.0 Tesla. Acad Radiol 2008;15(6):776–785.
12. Kirby M, Mathew L, Heydarian M, Etemad-Rezai R, McCormack DG, Parraga G. Chronic obstructive pulmonary disease: quantification of bronchodilator effects by using hyperpolarized ^{3}He MR imaging. Radiology 2011;261(1):283–292.
13. Kirby M, Kanhere N, Etemad-Rezai R, McCormack DG, Parraga G. Hyperpolarized helium-3 magnetic resonance imaging of chronic obstructive pulmonary disease exacerbation. J Magn Reson Imaging 2013;37(5):1223–1227.
14. Kirby M, Heydarian M, Svenningsen S, et al. Hyperpolarized 3He magnetic resonance functional imaging semiautomated segmentation. Acad Radiol 2012;19(2):141–152.
15. Kirby M, Pike D, Coxson HO, McCormack DG, Parraga G. Hyperpolarized (3) He ventilation defects used to predict pulmonary exacerbations in mild to moderate chronic obstructive pulmonary disease. Radiology 2014;273(3):887–896.
16. Kirby M, Mathew L, Wheatley A, Santyr GE, McCormack DG, Parraga G. Chronic obstructive pulmonary disease: longitudinal hyperpolarized (3)He MR imaging. Radiology 2010;256(1):280–289.
17. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images are more than pictures, they are data. Radiology 2016;278(2):563–577.
18. Han F, Wang H, Zhang G, et al. Texture feature analysis for computer-aided diagnosis on pulmonary nodules. J Digit Imaging 2015;28(1):99–115.
19. Kirby M, Pike D, McCormack DG, Lam S, Coxson HO, Parraga G. Longitudinal Computed Tomography and Magnetic Resonance Imaging of COPD: Thoracic Imaging Network of Canada (TINCan) Study Objectives. Chronic Obstr Pulm Dis (Miami) 2014;1(2):200–211.
20. Kirby M, Eddy RL, Pike D, et al. MRI ventilation abnormalities predict quality-of-life and lung function changes in mild-to-moderate COPD: longitudinal TINCan study. Thorax 2017;72(5):475–477.
21. Miller MR, Hankinson J, Brusasco V, et al. Standardisation of spirometry. Eur Respir J 2005;26(2):319–338.
22. Jones PW, Quirk FH, Baveystock CM, Littlejohns P. A self-complete measure of health status for chronic airflow limitation. The St. George's Respiratory Questionnaire. Am Rev Respir Dis 1992;145(6):1321–1327.
23. Heinrich MP, Jenkinson M, Bhushan M, et al. MIND: modality independent neighbourhood descriptor for multi-modal deformable registration. Med Image Anal 2012;16(7):1423–1435.
24. Haralick RM, Shanmugam K, Dinstein I. Textural features for image classification. IEEE Trans Syst Man Cybern 1973;SMC-3(6):610–621.
25. Tang X. Texture information in run-length matrices. IEEE Trans Image Process 1998;7(11):1602–1609.
26. Mishima M, Hirai T, Itoh H, et al. Complexity of terminal airspace geometry assessed by lung computed tomography in normal subjects and patients with chronic obstructive pulmonary disease. Proc Natl Acad Sci U S A 1999;96(16):8829–8834.
27. Capaldi DP, Zha N, Guo F, et al. Pulmonary Imaging Biomarkers of Gas Trapping and Emphysema in COPD: (3)He MR Imaging and CT Parametric Response Maps. Radiology 2016;279(2):597–608.
28. Park YS, Seo JB, Kim N, et al. Texture-based quantification of pulmonary emphysema on high-resolution computed tomography: comparison with density-based quantification and correlation with pulmonary function test. Invest Radiol 2008;43(6):395–402.
29. Xu Y, Sonka M, McLennan G, Guo J, Hoffman EA. MDCT-based 3-D texture classification of emphysema and early smoking related lung pathologies. IEEE Trans Med Imaging 2006;25(4):464–475.
30. Ginsburg SB, Lynch DA, Bowler RP, Schroeder JD. Automated texture-based quantification of centrilobular nodularity and centrilobular emphysema in chest CT images. Acad Radiol 2012;19(10):1241–1251.
31. Yang J, Angelini ED, Balte PP, et al. Unsupervised Discovery of Spatially-Informed Lung Texture Patterns for Pulmonary Emphysema: The MESA COPD Study. Med Image Comput Comput Assist Interv 2017;10433:116–124.
32. Shafiq-Ul-Hassan M, Zhang GG, Latifi K, et al. Intrinsic dependencies of CT radiomic features on voxel size and number of gray levels. Med Phys 2017;44(3):1050–1062.
33. Yasaka K, Akai H, Mackin D, et al. Precision of quantitative computed tomography texture analysis using image filtering: A phantom study for scanner variability. Medicine (Baltimore) 2017;96(21):e6993.
34. Madani A, Van Muylem A, Gevenois PA. Pulmonary emphysema: effect of lung volume on objective quantification at thin-section CT. Radiology 2010;257(1):260–268.
35. Schroeder JD, McKenzie AS, Zach JA, et al. Relationships between airflow obstruction and quantitative CT measurements of emphysema, air trapping, and airways in subjects with and without chronic obstructive pulmonary disease. AJR Am J Roentgenol 2013;201(3):W460–W470.
36. Coxson HO, Dirksen A, Edwards LD, et al. The presence and progression of emphysema in COPD as determined by CT scanning and biomarker expression: a prospective analysis from the ECLIPSE study. Lancet Respir Med 2013;1(2):129–136.