Published in final edited form as:

J Magn Reson Imaging. 2019 October ; 50(4): 1169–1181. doi:10.1002/jmri.26734.

Deep Convolutional Neural Networks with Multi-plane Consensus Labeling for Lung Function Quantification Using UTE Proton MRI

Wei Zha, PhD¹, Sean B. Fain, PhD^{1,2,3}, Mark L. Schiebler, MD², Scott K. Nagle, MD, PhD^{1,2,4}, Fang Liu, PhD²

¹Department of Medical Physics, University of Wisconsin-Madison

²Department of Radiology, University of Wisconsin-Madison

³Department of Biomedical Engineering, University of Wisconsin-Madison

⁴Department of Pediatrics, University of Wisconsin-Madison

Abstract

BACKGROUND: Ultrashort echo time (UTE) proton MRI has gained popularity for assessing lung structure and function in pulmonary imaging; however, the development of fast biomarker extraction and regional quantification has lagged behind due to labor-intensive lung segmentation.

PURPOSE: To evaluate a deep learning (DL) approach for automated lung segmentation to extract image-based biomarkers from functional lung imaging using 3D radial UTE oxygen-enhanced (OE) MRI.

STUDY TYPE: Retrospective study aimed to evaluate a technical development.

POPULATION: 45 human subjects, including 16 healthy volunteers, 5 asthma and 24 patients with cystic fibrosis.

FIELD STRENGTH/SEQUENCE: 1.5T MRI, 3D radial UTE (TE=0.08ms) sequence.

ASSESSMENT: Two 3D radial UTE volumes were acquired sequentially under normoxic and hyperoxic conditions. Automated segmentation of the lungs using 2D convolutional encoder-decoder based deep learning method, and the subsequent functional quantification via adaptive K-means were compared to the results obtained from the reference method, supervised region growing.

STATISTICAL TESTS: Relative to the reference method, the performance of DL on volumetric quantification was assessed using Dice coefficient with 95% confidence interval (CI) for accuracy, two-sided Wilcoxon signed-rank test for computation time and Bland-Altman analysis on the functional measure derived from the OE images.

RESULTS: The DL method produced strong agreement with supervised region growing for the right (Dice: 0.97; 95% CI = [0.96, 0.97]; $P < 0.001$) and left lungs (Dice: 0.96; 95% CI = [0.96,

0.97]; $P < 0.001$). The DL method averaged 46 seconds to generate the automatic segmentations in contrast to 1.93 hours using the reference method ($P < 0.001$). Bland-Altman analysis showed non-significant inter-method differences of volumetric ($P = 0.12$) and functional measurements ($P = 0.34$) in the left and right lungs.

DATA CONCLUSION: Deep learning provides rapid, automated and robust lung segmentation for quantification of regional lung function using ultrashort echo time proton MRI.

Keywords

Magnetic resonance imaging; Lung; Hyperoxia; Asthma; Cystic Fibrosis; Deep Learning

Introduction

Magnetic resonance imaging (MRI) of the chest can provide regional assessment of lung ventilation by inhaling oxygen (O_2) (1–4), hyperpolarized noble gases (5–8), or fluorinated gas (9–11). Oxygen-enhanced (OE) MRI using a three-dimensional (3D) radial ultrashort echo time (UTE) sequence supports quantitative differentiation of diseased vs. healthy lungs using whole lung ventilation defect percent (VDP) (4). This method does not require specialized multi-nuclear hardware or expensive specialty gases, while providing full chest images of regional ventilation with isotropic spatial resolution. A recent study demonstrated that the UTE signal-intensity measurements of the lung are reproducible and significantly correlated with computed tomography (CT) measures of emphysema and pulmonary function testing in human subjects with chronic obstructive pulmonary disease (12). Moreover, UTE proton MRI has been recently shown to have favorable comparison to CT findings of lung structural changes in subjects with evidence of cystic fibrosis (CF) (13–15).

Despite these rapid advances in pulmonary structural and functional imaging using UTE MRI, the development of a fast, reproducible and robust quantification tool for extracting potential biomarkers and regional image features remain to be developed (16). Segmentation of lung parenchyma from proton MRI is challenging due to modality-specific complexities including coil inhomogeneity, arbitrary intensity values, local magnetic susceptibility and the reduced proton density due to the large fraction of air space in the normal lung. The most recent state-of-the-art advances in lung segmentation methods for proton MRI have been focused on conventional echo time (TE) acquisitions, including 3D region growing method (17) and atlas-based estimation (18), which demand either labor-intensive manual correction or a high computational burden. These limitations are problematic in the presence of altered lung structure due to obstructive or restrictive pulmonary diseases, especially for 3D UTE proton images which are acquired at an isotropic spatial resolution.

Deep learning (DL) methods using convolutional neural networks (CNNs) have been successfully applied for tissue classification on chest CT images (19–21) and classification of pulmonary tuberculosis from chest radiographs (22). The deep learning methods have reportedly shown efficacy and efficiency on automated tissue classification for medical imaging using MRI, such as multi-class cartilage segmentation (23–25) and brain segmentation on nonhuman primates (26). However, these methods have not been applied to UTE proton MRI using O_2 to facilitate fast quantification of lung ventilation.

Our primary purpose was to evaluate a DL framework for automated lung segmentation from functional lung imaging using OE UTE proton MRI to support efficient functional quantification. Secondly, to understand disease-related structural alterations, the parenchymal signal-intensity assessments in the upper, middle, and lower lung regions automatically separated from the whole lung mask were compared in diseased vs normal groups. We hypothesized that a DL method would provide a more computationally efficient and robust alternative for volumetric and functional measurements from OE UTE proton MRI relative to the reference method of supervised region growing.

Materials and Methods

Human Subjects

Of three previous studies conducted on a GE 1.5T scanner (Signa HDx, GE Healthcare, Waukesha, WI), 45 OE MRI studies were collected from 45 unique subjects, including 16 healthy normal (age = 29.9 ± 11.0 year, 8/8 male/female split), 5 asthma (age = 43.2 ± 16.7 year, 4/1 male/female split), and 24 CF (age = 23.8 ± 10.5 year, 13/11 male/female split). All subjects gave informed consent, and all studies were carried out in accordance with the regulations of Internal Review Board and Health Insurance Portability and Accountability Act. Subject demographics are summarized in Table 1.

Asthma subjects were recruited through exploratory MRI substudy, Severe Asthma Research Program (SARP III) (27). Subjects were excluded if there was evidence of current infection, dependence on supplemental oxygen, or history of restrictive lung disease or cardiac disease. The inclusion criteria for the CF MRI protocol were an existing clinical diagnosis of CF, 10 years of age or older, and ability to return for study visits. Subjects were excluded for ventilator or oxygen dependence, history of lung transplant, contraindication to MRI, treatment with intravenous antibiotics for pulmonary exacerbation within 4 weeks of study visit and/or pregnancy. Supplemental Figure S1 enumerates the detailed inclusion and exclusion criteria for each subject group.

The whole lung VDP for 18 subjects (7 normal, 5 asthma and 6 CF) was previously reported as part of an interim study to develop a post-processing workflow for VDP quantification using OE MRI and compare VDP in diseased vs. normal groups (4). A comparison study on whole lung VDP measured from hyperpolarized ^3He vs. OE MRI for 24 subjects with CF was previously published (28).

3D Radial UTE Oxygen-enhanced MRI

Pulse sequence.—The 3D radial UTE sequence (29) was optimized to minimize TE and improve the signal-to-noise ratio (SNR) and image quality of lung imaging. The acquisition protocol used an axial slab-selective radiofrequency excitation with limited field of view (FOV), variable density readout gradients with eddy-current correction and oversampling in the readout direction by a factor of two to reduce artifact from excited tissues outside the nominal FOV. The flip angle was optimized to maximize contrast-to-noise ratio given the expected normoxic vs. hyperoxic differences in the longitudinal relaxation time (T1) (30). The 3D radial center-out trajectories provide full chest coverage in isotropic spatial

resolution and are typically insensitive to cardiac motion (31) due to the signal averaging from the acquisition of the center of k -space every repetition time (TR).

Experimental setup.—Each subject was placed supine in the MRI scanner, wearing a non-rebreather face mask (Supplemental Figure S2). The inlet of the face mask was connected to 3 sets of 7-foot-long medical grade oxygen tubing lines sequentially, leading out of the scanner room through the waveguide and to the outlet port of the oxygen mixer. The two inlets of the oxygen mixer were connected to the oxygen tanks which contain medical grade room air (21% O₂, normoxic) and 100% O₂ (hyperoxic) respectively, allowing the inspired gas to switch in the range of 21% and 100% O₂ at the chosen flow rate of 15 L/min throughout the scan session.

MR examination.—The timeline of the imaging protocol is illustrated in the supplemental Figure S2. Subjects were coached to maintain tidal breathing during the approximately 9-minute MRI exam. First, each subject breathed 21% O₂ for the first 3.5-minute UTE acquisition. Second, the O₂ concentration was switched to 100%. After 2 minutes of tidal breathing at 100% O₂ to avoid the transient effects of gas wash-in, the subject then underwent a second 3.5-min UTE acquisition under hyperoxic free-breathing. Prospective respiratory gating to end-expiration was applied with adaptive feedback from the respiratory bellows signal to gauge a 50% acceptance window (illustrated in the supplemental Figure S2), leading to an acquisition at approximately the lung inflation volume corresponding to functional residual capacity (FRC). The UTE scan efficiency was 50% due to the 50% acceptance window and the fixed number of radial projections. Additional details regarding the acquisition and reconstruction have been previously published (29, 30) and summarized in the Supplementary Material [Online].

Regional Ventilation Quantification for OE MRI

A previously developed workflow including data preprocessing, deformable registration and retrospective lung density correction was used to generate PSE maps from the normoxic and hyperoxic UTE volumes acquired for each subject (4). The normoxic and hyperoxic UTE volumes were preprocessed automatically with image denoising, intensity correction using bias field inhomogeneity estimation (32) and intensity normalization to the intensity range [0, 1]. These UTE volumes were then cropped semi-automatically in axial, coronal and sagittal planes in order to reduce the computational cost in image registration. This data cropping involved ~1-minute manual interaction per dataset to ensure the entire lung was preserved in the cropped data. After registration and density correction, the ventilation-weighted PSE map was computed as $PSE = (S_{100\%} - S_{21\%})/S_{21\%} \cdot 100\%$, where $S_{21\%}$ and $S_{100\%}$ represent the signal intensity of the normoxic and hyperoxic UTE images respectively. The low-intensity regions from the PSE map was considered as ventilation defects, which were quantified automatically as VDP by a machine learning based approach, adaptive K -means (4, 33), using the binary lung mask segmented using either the reference method or the DL approach. Lung segmentation and ventilation defect quantification was performed individually by an imaging scientist (W.Z.) with 11 years of experience in cardiothoracic MRI who was blinded to clinical information under the supervision of a cardiothoracic

radiologist (S.K.N.) with 10 years of experience. Details of the regional ventilation quantification algorithm are presented in the Supplementary Material [Online].

Reference Standard Lung Segmentation Using 3D Region Growing

The lungs were segmented from 45 preprocessed hyperoxic UTEs individually in the axial slices using 3D region growing (17) with supervised correction by an imaging scientist (W.Z.). This 3D region growing method was proposed as an automatic lung segmentation tool for perfusion studies using three types of pulmonary proton MRI acquired at conventional TE (17). As illustrated in Figure 1, the preprocessed 3D image volume was downsampled for a fast, automated selection of the lung seed points based on a Euclidean distance transformation: these seed points were used for the 3D region growing with empirical thresholds to obtain a coarse 3D lung mask. Several supervised morphometric correction steps were required for an accurate delineation of lung parenchyma due to the improved visualization of disease-related structural changes on UTE relative to more conventional methods (15) and apical MR signal dropout due to field inhomogeneity. These binary lung masks after supervised correction were used as the reference standard to train and evaluate the DL network models.

Models Architecture for Automated Lung Segmentation

The DL framework for lung segmentation (Figure 2, A) used a two-dimensional (2D) convolutional encoder-decoder (CED) architecture (34), which has been successfully applied for cartilage and brain tissue segmentation (23, 26, 35, 36). The encoder network uses the same 13 Visual Geometry Group 16 convolutional layers and the decoder uses a mirrored structure of the encoder network with max-pooling replaced by up-sampling process. A symmetric shortcut connection (SC) between the encoder and decoder network is added to enhance the segmentation performance. (37). A multi-plane consensus labeling strategy (38) (Figure 2, B) was implemented to allow an efficient 2D CNN framework while integrating 3D isotropic image features (39).

Training and Evaluation

The 45-subject hyperoxic UTE data were stratified into 5 sets, resulting in 9 hyperoxic UTE in each set. A five-fold cross-validation (Figure 3) was performed in order to ensure that the data used in testing phases were unseen in the training phases. This cross-validation procedure allows to test the proposed DL method on the hyperoxic UTE data from all 45 subjects by combining the predicted lung masks from 5 sets of hyperoxic testing data which are unseen from the corresponding model training.

In the training phase, the network was trained individually in three orthogonal (axial, coronal and sagittal) planes. The axial training used random initial network weights for the CED with a total iteration steps corresponding to 20 epochs for training convergence. For the subsequent training in the coronal and sagittal planes, transfer learning was applied for better initialization of network parameters and computational efficiency of the training process (40, 41). Specifically, the weights resulted from axial training was used as the initial weights for coronal training; and the weights from the coronal training was used as the

initial weights for the sagittal weights. Consequently, the coronal and sagittal trainings were carried out with a total iteration steps corresponding to 10 epochs for training convergence.

Each single-plane training process was formulated as an optimization problem to optimize the network parameters by minimizing the difference between the network's output and the corresponding reference mask labels using cross-entropy loss (42) as the cost function. The well-trained network model was selected from the iteration at which the loss computed using the validation data is the lowest among all the iterations. In the testing phase, the well-trained network model resulting from each single-plane training is used as a front-end segmentation classifier to produce a volumetric segmentation of the lungs in a slice-by-slice fashion, which is referred as a single-plane predicted mask. For each UTE, the three sets of orthogonal, binary single-plane predicted masks were combined within the 3D context using consensus labeling to obtain the multi-plane consensus masks, or *DL predicted masks*. The hardware and software configurations for training and testing are provided in the Supplementary Material [Online].

To evaluate parenchymal signal intensity, the well-trained network models were applied for the normoxic UTE data acquired prior to the hyperoxic UTE from OE MRI, similar to the strategy for longitudinal precision assessment (25). These normoxic UTE data have no reference masks available.

Statistical Analysis

The performance of lung segmentation using DL was evaluated in three respects. First, accuracy performance of the CNN model were gauged by computing the Dice coefficient $(\frac{2|T \cap P|}{|T| + |P|})$ (18) for all 45 hyperoxic lung volumes, where T is the reference mask and P is the predicted mask. One-way analysis of variance (ANOVA) was used to perform group-wise comparisons of the Dice coefficient among the reference masks, DL predicted masks, and single-plane predicted masks. Homogeneity of variance was tested using Levene's test. The model adequacy was checked, and Welch test was used to assess global effect of ANOVA if violated (43). When the global test confirmed the differences were significant, the Tukey-Kramer procedure was performed to adjust the p values of all pair-wise differences and control the pairwise comparisons among the groups jointly in order to avoid type I error rate inflation (44, 45). The computation cost of DL vs. reference method was compared using Wilcoxon signed-rank test.

Second, Bland-Altman analysis with 95% limits of agreement (LOA) was used to compare the VDP measurements of the right and left lungs which were quantified using the DL predicted masks and the reference masks, respectively. The inter-method agreement was assessed using nonparametric, two-sided Wilcoxon signed-rank test, if the normality assumption was not valid.

Third, the DL predicted masks of the normoxic UTE volumes were divided along the superior/inferior direction into three equal volumes: upper, middle, and lower regions in axial view. The normalized normoxic UTE data from each subject group were grouped according to comparable acquisition parameters. The regional (upper, middle and lower)

lung parenchymal signal intensity in normal vs. CF (3.2mm isotropic resolution) and normal vs. asthma (1.25mm isotropic resolution) were compared using Wilcoxon signed-rank test. All statistical analyses were performed using SAS 9.4 (SAS Institute, Cary, NC). A $P < 0.05$ was considered statistically significant.

Results

Functional and demographic data are summarized in Table 1. The whole lung VDP from OE MRI was significantly higher in the diseased vs. normal groups (ANOVA $P < 0.001$), suggesting this study cohort contains patients with significant loss of lung function. Of all 45 subjects aged 11–59 years, ten subjects with CF were 18 years old.

Deep Learning Performances on Lung Segmentation

The performance of the DL method was evaluated on the hyperoxic UTE volumes from 45 subjects. It took 1.93 ± 1.67 hours to obtain a reference mask, depending on the existence and the severity of structural abnormalities. In contrast, the DL method with consensus labeling framework took ~46 seconds to obtain the DL predicted mask (multi-plane consensus mask) per UTE volume, suggesting significantly ($P < 0.001$) reduced computation time using DL vs. the supervised region growing.

Figure 4 shows examples of segmented lung parenchyma using DL compared to the reference standard in a subject with CF. Relative to the single-plane predicted masks, the multi-plane consensus mask removed small errors and conformed visually better to the lung periphery. The Dice coefficients calculated between the reference and the multi-plane consensus masks were 0.97 (95% CI = [0.96, 0.97], $P < 0.0001$) for the right lung and 0.96 (95% CI = [0.96, 0.97], $P < 0.0001$) for the left lung (Table 2). Compared to the single-plane predicted masks, the multiple comparisons found the multi-plane consensus mask had significantly better (ANOVA Welch's test $P = 0.03$) Dice coefficients compared to the coronal-plane predicted masks for the left lung. For the right lung, no statistical differences ($P = 0.15$) were found among single-plane predicted masks and multi-plane consensus masks for the right lung.

In three different subjects with varying physiological conditions (Supplemental Figure S3), the DL predicted masks demonstrate overall similar delineation of lung parenchyma relative to the reference masks.

Quantitative Measurements of Lung Volume and Function

The right volume ranged 0.80–2.56 L and the left lung volume ranged 0.56–2.21 L. A breakdown of lung volumes and VDP measurements of the right and left lungs assessed using the DL predicted masks vs. reference masks in each subject group showed no significant differences with $P \geq 0.062$ (Table 3). Bland-Altman analysis (Figure 5a and 5c) showed that the average inter-method absolute volume differences were non-significant: -0.018 L ($P = 0.12$) with 95% LOA of $[-0.13, 0.09]$ L for the right lung and -0.013 L ($P = 0.24$) with 95% LOA of $[-0.12, 0.09]$ L for the left lung.

For VDP, Bland-Altman analysis (Figure 5b and 5d) showed good agreement on the measurements extracted from the DL predicted masks and the reference masks for the right lung (-0.061% with 95% LOA = $[-3.10\%, 2.98\%]$ and $P=0.34$) and the left lung (-0.14% with 95% LOA = $[-2.72\%, 2.45\%]$ and $P=0.81$).

Normoxic UTE Parenchymal Signal-intensity Alterations

In the upper, middle and lower regions of the lungs on the normalized normoxic UTE images, the average parenchymal signal intensity for CF was significantly higher than for normal subjects in each of the lung regions with $P<0.0001$ (Figure 6). This may attribute to the structural changes evidenced in CF, including consolidations, varicoid bronchiectasis and mucus plugging (Figure 7). The parenchymal signal intensity for asthma was not statistically different from the normal subjects in all three lung regions with $P=0.44$, likely attributed to smaller sample size in asthma which limited power for comparison.

Discussion

Our study demonstrates the speed, capability and efficiency of a deep CNN approach for automated lung segmentation of UTE proton MRI. The finding of significantly higher regional parenchymal signal intensity and higher whole lung VDP in CF vs. normal subjects indicates that the CF group contains subjects with significant lung disease. The excellent Bland-Altman analysis agreement of the volumetric and functional measurements extracted from DL vs. reference segmentation shows the robustness of the DL method. Given the substantial equivalence of the segmentation performance, DL has the distinct advantage over 3D region growing in its significant reduction in computational cost (46 seconds for DL vs. 1.9 hours for region growing, $P<0.001$). We found that OE UTE proton MRI ventilation imaging post processing with DL technique provides a rapid and robust method for the extraction of functional biomarkers with complementary structural information to improve the understanding of regional pulmonary disease and possibly the longitudinal rate of progression in vulnerable subjects without the use of medical radiation.

The evidence of structural changes on UTE proton MRI of the lung greatly increased the computational cost and the uncertainty for an accurate and objective lung extraction using the supervised region growing. The proposed DL approach achieved strong Dice coefficients using 45 subjects' pulmonary UTE OE MRI data that encompassed different lung volumes ranging from 1.36L to 4.77L. The equivalent performance of lung segmentation suggests the distinct advantage of DL over 3D region growing in its significant reduction in computational cost, which is a common trait required for clinical research. Aside from the speed and accuracy performance, the proposed DL method has a distinct advantage over the state-of-the-art methods in that it provides an end-to-end feature-based solution with no parameter tuning due to varying SNR (i.e., region growing thresholds) and subject body size (i.e., gradient step size used in registration).

Given the wide range of age and disease severity represented in the CF cohort of the present study, the DL approach performs well in the setting of significant, potentially confounding lung pathology. It is noteworthy that the UTE data of this feasibility study available from OE MRI have lower native spatial resolution relative to the submillimeter resolution UTE (14,

15) which has demonstrated CT-like structure findings. The performance of DL for the UTE mostly at 3.2mm native resolution in the present study implies that the DL approach would perform better on structural UTE data with higher spatial resolution (smaller voxel size).

Three-dimensional radial UTE proton MRI can depict the pulmonary parenchymal structural changes of CF in a similar way to CT (14, 16). A recent study demonstrates that the UTE signal intensity of the lungs is reproducible and significantly correlated to CT measures of emphysema and PFTs in chronic obstructive pulmonary disease (12). In a healthy cohort, the changes in T2 and T2* relaxation times in the lungs under hypoxia vs. normoxia were reported using a 3D multiecho inversion recovery UTE sequence (46). A feasibility study demonstrated pulmonary perfusion imaging using 3D radial UTE sequence (47). While these recent technical advances make 3D UTE MRI a promising tool for pulmonary functional and structural imaging, the quantitative biomarkers that could easily and reliably be extracted from the images can potentially drive these UTE techniques towards clinical research. Therefore, a deep CNN approach is critical for a fast extraction of quantitative measures in the lungs.

There are several limitations to this work. First, this retrospective feasibility study has a small sample size. Future work will recruit more study subjects, especially patients with asthma, and evaluate the model architecture using a hold-out test set when more data become available. Second, while our neural network prediction shows promising performance advantages in comparison to the reference standard, there is still room for improvement in accuracy. For example, setting up the training for a specific subject cohort such as CF; using Dice or Jaccard coefficient as the loss function in the training for segmentation. Third, the retrospective data have varying SNR (Supplemental Figure S3) due to the three different native spatial resolutions used in acquisitions. It is worth further investigation on the model performance for data with better consistency. Additionally, the reference masks in the present study were generated by a single imaging scientist with limited consensus reading by a cardiothoracic radiologist (***) due to the labor-intensive manual correction. Future work will look into inter-reader variability, and investigate the uncertainty of the network model by incorporating probabilistic CED such as Bayesian CNN (26) to generate the confidence map for assessing the segmentation repeatability and robustness. The preliminary results of this study show insight into the application of DL in the field of pulmonary proton MRI.

In conclusion, these preliminary results show the potential role of deep learning for a rapid, automated and robust extraction of quantitative biomarkers from functional lung imaging using oxygen-enhanced ultrashort echo time proton MRI. Further developments in deep learning may improve the clinical value of oxygen-enhanced pulmonary UTE proton MRI for structural and functional assessments without ionizing radiation.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Funding Acknowledgements:

NIH/NHLBI R01 HL126771, NIH/NCATS S10 OD016394

Abbreviations:

CF	cystic fibrosis
CNN	convolutional neural network
DL	deep learning
OE	oxygen-enhanced
UTE	ultrashort echo time
VDP	ventilation defect percent

References

1. Edelman RR, Hatabu H, Tadamura E, Li W, Prasad P V. Noninvasive assessment of regional ventilation in the human lung using oxygen-enhanced magnetic resonance imaging. *Nat Med* 1996; 2:1236–1239. [PubMed: 8898751]
2. Ohno Y, Nishio M, Koyama H, et al. Asthma: comparison of dynamic oxygen-enhanced MR imaging and quantitative thin-section CT for evaluation of clinical treatment. *Radiology* 2014; 273:907–916. [PubMed: 25102370]
3. Zhang W-J, Niven RM, Young SS, Liu Y- Z, Parker GJM, Naish JH. Dynamic oxygen-enhanced magnetic resonance imaging of the lung in asthma -- initial experience. *Eur J Radiol* 2015; 84:318–326. [PubMed: 25467640]
4. Zha W, Kruger SJ, Johnson KM, et al. Pulmonary Ventilation Imaging in Asthma and Cystic Fibrosis Using Oxygen-Enhanced 3D Radial Ultrashort Echo Time MRI. *J Magn Reson Imaging* 2018; 47:1287–1297. [PubMed: 29086454]
5. Woodhouse N, Wild JM, Paley MNJ, et al. Combined helium-3/proton magnetic resonance imaging measurement of ventilated lung volumes in smokers compared to never-smokers. *J Magn Reson Imaging* 2005; 21:365–369. [PubMed: 15779032]
6. Fain SB, Gonzalez-Fernandez G, Peterson ET, et al. Evaluation of Structure-Function Relationships in Asthma using Multidetector CT and Hyperpolarized He-3 MRI. *Acad Radiol* 2008; 15:753–762. [PubMed: 18486011]
7. Pike D, Kirby M, Eddy RL, et al. Regional Heterogeneity of Chronic Obstructive Pulmonary Disease Phenotypes: Pulmonary 3He Magnetic Resonance Imaging and Computed Tomography. *COPD* 2016:1–9.
8. Marshall H, Horsley A, Taylor CJ, et al. Detection of Early Subclinical Lung Disease in Children with Cystic Fibrosis by Lung Ventilation Imaging with Hyperpolarised Gas MRI. *Thorax* 2017:760–762. [PubMed: 28265032]
9. Halawish AF, Moon RE, Foster WM, et al. Perfluoropropane gas as a magnetic resonance lung imaging contrast agent in humans. *Chest* 2013; 144:1300–1310. [PubMed: 23722696]
10. Couch MJ, Ball IK, Li T, et al. Inert fluorinated gas MRI: a new pulmonary imaging modality. *NMR Biomed* 2014; 27:1525–1534. [PubMed: 25066661]
11. Gutberlet M, Kaireit TF, Voskresenzyev A, et al. Free-breathing Dynamic (19)F Gas MR Imaging for Mapping of Regional Lung Ventilation in Patients with COPD. *Radiology* 2018; 286:1040–1051. [PubMed: 28972817]
12. Ma W, Sheikh K, Svenningsen S, et al. Ultra-short Echo-time Pulmonary MRI: Evaluation and Reproducibility in COPD Subjects with and without Bronchiectasis. *J Magn Reson Imaging* 2015; 41:1465–1474. [PubMed: 24965907]

13. Roach DJ, Cremillieux Y, Fleck RJ, et al. Ultrashort Echo-Time Magnetic Resonance Imaging Is a Sensitive Method for the Evaluation of Early Cystic Fibrosis Lung Disease. *Ann Am Thorac Soc* 2016; 13:1923–1931. [PubMed: 27551814]
14. Dourmes G, Menut F, Macey J, et al. Lung Morphology Assessment of Cystic Fibrosis Using MRI with Ultra-short Echo Time at Submillimeter Spatial Resolution. *Eur Radiol* 2016; 26:3811–3820. [PubMed: 26843010]
15. Nagle SK, Francois CJ, Poranski M, Bell LC, Johnson KM, Fain SB. Conventional and Ultrashort Echo Time MRI Compared with CT. In *North Am Cyst Fibros Conf Orlando, FL*; 2016.
16. Nagle SK, Puderbach M, Eichinger M, Altes TA. Magnetic Resonance Imaging of the Lung: Cystic Fibrosis In *Med Radiol*. Springer, Berlin, Heidelberg; 2017:1–15.
17. Kohlmann P, Strehlow J, Jobst B, et al. Automatic lung segmentation method for MRI-based lung perfusion studies of patients with chronic obstructive pulmonary disease. *Int J Comput Assist Radiol Surg* 2015; 10:403–417. [PubMed: 24989967]
18. Tustison NJ, Qing K, Wang C, Altes T a., Mugler JP Atlas-based estimation of lung and lobar anatomy in proton MRI. *Magn Reson Med* 2015; 00:n/a-n/a.
19. Microbiana B, Hidalgo D, Anthimopoulos M, et al. Lung Pattern Classification for Interstitial Lung Diseases Using a Deep Convolutional Neural Network. *IEEE Trans Med Imaging* 2016; 35:1207–1216. [PubMed: 26955021]
20. Gao M, Bagci U, Lu L, et al. Holistic classification of CT attenuation patterns for interstitial lung diseases via deep convolutional neural networks. *Comput methods Biomech Biomed Eng Imaging Vis* 2018; 6:1–6. [PubMed: 29623248]
21. Arindra A, Setio A, Ciompi F, et al. Pulmonary Nodule Detection in CT Images: False Positive Reduction Using Multi-View Convolutional Networks. *IEEE Trans Med Imaging* 2016; 35:1160–1169. [PubMed: 26955024]
22. Lakhani P, Sundaram B. Deep Learning at Chest Radiography: Automated Classification of Pulmonary Tuberculosis by Using Convolutional Neural Networks. *Radiology* 2017; 284:574–582. [PubMed: 28436741]
23. Liu F, Zhou Z, Jang H, Samsonov A, Zhao G, Kijowski R. Deep convolutional neural network and 3D deformable approach for tissue segmentation in musculoskeletal magnetic resonance imaging. *Magn Reson Med* 2018; 79:2379–2391. [PubMed: 28733975]
24. Zhou Z, Zhao G, Kijowski R, Liu F. Deep convolutional neural network for segmentation of knee joint anatomy. *Magn Reson Med* 2018.
25. Norman B, Pedoia V, Majumdar S. Use of 2D U-Net Convolutional Neural Networks for Automated Cartilage and Meniscus Segmentation of Knee MR Imaging Data to Determine Relaxometry and Morphometry. *Radiology* 2018; 288:177–185. [PubMed: 29584598]
26. Zhao G, Liu F, Oler JA, Meyerand ME, Kalin NH, Birn RM. Bayesian convolutional neural network based MRI brain extraction on nonhuman primates. *Neuroimage* 2018; 175:32–44. [PubMed: 29604454]
27. Jarjour NN, Erzurum SC, Bleecker ER, et al. Severe asthma: Lessons learned from the National Heart, Lung, and Blood Institute Severe Asthma Research Program. *Am J Respir Crit Care Med* 2012:356–362. [PubMed: 22095547]
28. Zha W, Nagle SK, Cadman R V, Schiebler ML, Fain SB. Three-dimensional Isotropic Functional Imaging of Cystic Fibrosis Using Oxygen-enhanced MRI: Comparison with Hyperpolarized (3)He MRI. *Radiology* 2019; 290:229–37. [PubMed: 30351258]
29. Johnson KM, Fain SB, Schiebler ML, Nagle S. Optimized 3D Ultrashort Echo Time Pulmonary MRI. *Magn Reson Med* 2013; 70:1241–1250. [PubMed: 23213020]
30. Kruger SJ, Fain SB, Johnson KM, Cadman R V, Nagle SK. Oxygen-enhanced 3D Radial Ultrashort Echo Time Magnetic Resonance Imaging in the Healthy Human Lung. *NMR Biomed* 2014; 27:1535–1541. [PubMed: 24984695]
31. Stehning C, Bornert P, Nehrke K, Eggers H, Dossel O. Fast isotropic volumetric coronary MR angiography using free-breathing 3D radial balanced FFE acquisition. *Magn Reson Med* 2004; 52:197–203. [PubMed: 15236387]
32. Tustison NJ, Avants BB, Cook PA, et al. N4ITK: Improved N3 bias correction. *IEEE Trans Med Imaging* 2010; 29:1310–20. [PubMed: 20378467]

33. Zha W, Niles DJ, Kruger SJ, et al. Semiautomated Ventilation Defect Quantification in Exercise-induced Bronchoconstriction Using Hyperpolarized Helium-3 Magnetic Resonance Imaging: A Repeatability Study. *Acad Radiol* 2016; 23:1104–1114. [PubMed: 27263987]
34. Badrinarayanan V, Kendall A, Cipolla R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Scene Segmentation. *IEEE Trans Pattern Anal Mach Intell* 2017.
35. Liu F, Jang H, Kijowski R, Bradshaw T, McMillan AB. Deep Learning MR Imaging-based Attenuation Correction for PET/MR Imaging. *Radiology* 2018; 286:676–684. [PubMed: 28925823]
36. Liu F, Zhou Z, Samsonov A, et al. Deep Learning Approach for Evaluating Knee MR Images: Achieving High Diagnostic Performance for Cartilage Lesion Detection. *Radiology* 2018:172986.
37. Brosch T, Tang LYW, Yoo Y, Li DKB, Traboulsee A, Tam R. Deep 3D Convolutional Encoder Networks With Shortcuts for Multiscale Feature Integration Applied to Multiple Sclerosis Lesion Segmentation. *IEEE Trans Med Imaging* 2016; 35:1229–1239. [PubMed: 26886978]
38. Zha W, Fain SB, Schiebler ML, Nagle SK, Liu F. Improved Segmentation of Proton MRI Lung Volume Using a 2.5D Deep Convolutional Neural Network ISMRM Work Mach Learn Part I, Pacific Grove, CA, USA 2018.
39. Prasoon A, Petersen K, Igel C, Lauze F, Dam E, Nielsen M. Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network. *Med Image Comput Comput Assist Interv* 2013; 16(Pt 2):246–253. [PubMed: 24579147]
40. Shin HC, Roth HR, Gao M, et al. Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning. *IEEE Trans Med Imaging* 2016; 35:1285–1298. [PubMed: 26886976]
41. Tajbakhsh N, Shin JY, Gurudu SR, et al. Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning? 2017; 35:1299–1312.
42. Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. *Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit* 2015; 07–12–June:3431–3440.
43. Kohr RL, Games PA. Robustness of the analysis of variance, the welch procedure and a box procedure to heterogeneous variances. *J Exp Educ* 1974; 43:61–69.
44. Schiros CG, Dell'Italia LJ, Gladden JD, et al. Magnetic resonance imaging with 3-dimensional analysis of left ventricular remodeling in isolated mitral regurgitation: implications beyond dimensions. *Circulation* 2012; 125:2334–2342. [PubMed: 22496130]
45. Zha W, Schiros CG, Reddy G, et al. Improved Right Ventricular Performance with Increased Tricuspid Annular Excursion in Athlete's Heart. *Front Cardiovasc Med* 2015; 2.
46. Triphan SMF, Breuer FA, Gensler D, Kauczor HU, Jakob PM. Oxygen enhanced lung MRI by simultaneous measurement of T1 and T2* during free breathing using ultrashort TE. *J Magn Reson Imaging* 2015; 41:1708–1714. [PubMed: 25044618]
47. Bauman G, Johnson KM, Bell LC, et al. Three-dimensional pulmonary perfusion MRI with radial ultrashort echo time and spatial-temporal constrained reconstruction. *Magn Reson Med* 2015; 73:555–564. [PubMed: 24604452]

Figure 1.

Workflow of supervised 3D region growing. a. The hyperoxic UTE volume is preprocessed with de-noising, intensity correction and normalization, and smoothing. b. The 3D volume is subsampled for a fast estimation of body mask by thresholding after the histogram analysis. c. In the mid-lung image slice, the lung mask is estimated using K-means clustering. d. Boundaries of the estimated mask (c) is taken as input for a Euclidean distance transformation. e. Points with largest distance to the lung boundaries are automatically selected as seed points. f. Coarse 3D lung mask is resulted from 3D region growing using the seed points (e). (g) Final reference lung mask is determined by morphometric and manual correction for apical MR signal dropout and high-intensity structural abnormalities.

Figure 2.

A, Illustration of the convolutional encoder-decoder (CED) architecture features a connected encoder and decoder network. The encoder network uses the same 13 Visual Geometry Group 16 (VGG16) convolutional layers and the decoder uses a mirrored structure of the encoder network with max-pooling replaced by upsampling process. A symmetric shortcut connection (SC) between the encoder and decoder network is added to enhance the segmentation performance. *B*, Flowchart shows the training and testing phases of the multi-plane consensus labeling framework for lung segmentation of UTE proton image volumes at isotropic spatial resolution. 2D = two-dimensional, BN = batch normalization, ReLU = rectified-linear activation.

Figure 3.

Data flow and process. All 45 pairs of normoxic and hyperoxic UTE lung volumes acquired for oxygen-enhanced MRI were pre-processed. The 45 hyperoxic UTEs were randomly stratified into 5 non-overlapping data sets, with each set containing 9 UTE volumes. A five-fold cross-validation was performed to ensure the training and testing phases used independent sets of data by repeating the training and testing phases 5 times as follows. Each time, five data sets were divided into training, validation and testing sets with a 32/4/9 split and were then broken down into their respective 2D sections to be used as inputs for a convolutional encoder-decoder model. Each of the five well-trained models was used to predict the lung masks for the corresponding set of hyperoxic UTE testing data. By combining the deep learning predicted masks from 5 sets of hyperoxic testing data, the performance of deep learning was evaluated relative to the reference method on hyperoxic UTE volumes from all 45 participants. These deep learning predicted masks were then used for volumetric and functional quantification. The well-trained network models were applied

for the normoxic UTE data available from OE MRI for regional parenchymal signal-intensity assessment.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 4.

Examples of the segmented lung parenchyma using deep learning trained network models compared to the reference mask by supervised region growing for a 37-year-old female. The multi-plane consensus mask removed small errors outside of the lungs in sagittal- and coronal-plane predicted masks, conformed better to the lung peripheries in the proton images than that observed in single-plane (axial, coronal and sagittal) predicted mask, and corrected the errors stemming from consolidations in the sagittal-plane predicted mask and pleural findings in the axial-plane predicted mask.

Figure 5. Bland-Altman plots of lung volume and ventilation defect percent (VDP) for the right (a, b) and left (c, d) lungs suggest non-significant ($P = 0.12$) differences of the volumetric and functional measurements extracted from the deep learning predicted masks and the reference masks. A $p < 0.05$ was considered significant.

Figure 6.

Boxplots of the average parenchymal signal intensity in the upper, middle and lower regions of the lung from the normalized normoxic UTE images acquired at approximately functional residual capacity for a. normal vs. cystic fibrosis (CF) at 3.2mm native isotropic spatial resolution and b. normal vs. asthma acquired at 1.25mm isotropic native resolution. In each lung region, the average parenchymal signal intensity in CF (Upper: 0.090 ± 0.024 , $P < 0.001$; Middle: 0.096 ± 0.019 , $P < 0.001$; Lower: 0.089 ± 0.015 , $P < 0.001$) were significantly higher than that of the normal group (Upper: 0.037 ± 0.0085 ; Middle: 0.041 ± 0.0091 ; Lower:

0.039±0.0068). No significant differences were found in the regional comparisons between the normal and asthma. Boxes extend vertically between the 25th and 75th percentiles, the whiskers extend to the most extreme data that are not considered outliers, and outliers are plotted with plus sign. The central mark represents the median. The symbol * denotes significantly differences found vs. normal in each noted lung region. A $p<0.05$ was considered significant.

Figure 7.

Combined visualization of structural abnormalities of CF (arrows in white) identified from axial image slices of normoxic free-breathing ultrashort echo time (UTE) proton MRI for a 31-year-old female (a, b) and a 24-year-old female participant (c,d). UTE proton MRI visualizes structural abnormalities evidenced in cystic fibrosis, including pleural finding, mucus plugging, consolidation, sacculation and bronchiectasis. These high-intensity structural alterations may attribute to the higher parenchymal signal intensity seen in CF.

Table 1

Demographic breakdown by disease group for the 45 study subjects in the study cohort that underwent oxygen-enhanced UTE MRI

	Male	Female	Total	<i>P</i> value Male vs. Female
Normal	8	8	16	
Age (years)	31.1 ± 9.8 (22.9, 39.4)	28.8 ± 12.5 (18.3, 39.2)	29.9 ± 11.0 (24.1, 35.8)	0.40
OE MRI lung inflation volume (L)	2.79 ± 0.59 (2.29, 3.29)	2.07 ± 0.39 (1.74, 2.40)	2.43 ± 0.61 (2.10, 2.76)	0.028
OE MRI VDP (%)	3.09 ± 1.57 (1.77, 4.40)	5.10 ± 6.05 (0.05, 10.16)	4.10 ± 4.39 (1.76, 6.44)	0.38
Asthma	4	1	5	
Age (years)	40.75 ± 18.25 (11.72, 69.78)	53	43.20 ± 16.72 (22.43, 63.97)	>0.99
OE MRI lung inflation volume (L)	2.99 ± 0.90 (1.56, 4.42)	2.02	2.80 ± 0.89 (1.69, 3.90)	0.40
OE MRI VDP (%)	18.61 ± 11.93 (-0.37, 37.60)	32.06	21.30 ± 11.96 (6.46, 36.15)	0.40
Cystic Fibrosis	13	11	24	
Age (years)	27.08 ± 11.14 (20.35, 33.81)	19.82 ± 8.54 (14.08, 25.56)	23.75 ± 10.49 (19.32, 28.18)	0.087
OE MRI lung inflation volume (L)	3.11 ± 0.93 (2.55, 3.68)	2.21 ± 0.88 (1.62, 2.80)	2.70 ± 1.0 (2.28, 3.12)	0.0065
OE MRI VDP (%)	23.12 ± 15.74 (13.61, 32.63)	19.26 ± 12.78 (10.67, 27.84)	21.35 ± 14.29 (15.31, 27.38)	0.27

The functional measurement ventilation defect percent (VDP) from OE MRI was significantly higher in asthma vs. normal ($P=0.011$) and in cystic fibrosis vs. normal ($P<0.001$). OE MRI lung volume was acquired approximately at functional residual capacity. Data are presented as mean ± standard deviation (95% Confidence levels). A $p<0.05$ is considered statistically significant for Male vs. Female using two-sided Wilcoxon signed-rank test. (Abbreviations: OE—oxygen-enhanced; VDP— ventilation defect percent)

Table 2

Dice comparison of the lung masks segmented using deep learning and the reference method (N=45)

Lung	Axial-plane predicted masks	Coronal-plane predicted masks	Sagittal-plane predicted masks	Multi-plane consensus masks
Right	0.96 ± 0.021 (0.96, 0.97)	0.96 ± 0.022 (0.95, 0.97)	0.96 ± 0.015 (0.96, 0.97)	0.97 ± 0.015 (0.96, 0.97)
Left	0.96 ± 0.015 (0.95, 0.96)	0.95 ± 0.027* (0.94, 0.96)	0.95 ± 0.022 (0.95, 0.96)	0.96 ± 0.012 (0.96, 0.97)

Data are presented as mean ± standard deviation (95% Confidence levels). The Dice coefficients using each set single-plane predicted masks were compared to those using the multi-plane consensus masks. An asterisk symbol denotes a significant difference found between single-plane predicted masks vs. multi-plane consensus masks using one-way analysis of variance (ANOVA). A $p < 0.05$ is considered statistically significant.

Table 3

Comparisons of lung volumes and functional measure ventilation defect percent extracted from deep learning and the reference method

	Lung	Deep learning predicted masks	Reference masks	<i>P</i> value
Hyperoxic lung volume at functional residual capacity (L)				
Normal (N=16)	Right	1.33 ± 0.34 (1.15, 1.51)	1.34 ± 0.33 (1.16, 1.52)	0.74
	Left	1.09 ± 0.29 (0.94, 1.25)	1.09 ± 0.28 (0.94, 1.25)	0.86
Asthma (N=5)	Right	1.47 ± 0.44 (0.92, 2.02)	1.48 ± 0.45 (0.92, 2.04)	0.81
	Left	1.30 ± 0.40 (0.81, 1.79)	1.32 ± 0.43 (0.78, 1.86)	0.63
Cystic fibrosis (N=24)	Right	1.45 ± 0.53 (1.23, 1.67)	1.48 ± 0.54 (1.25, 1.71)	0.062
	Left	1.20 ± 0.45 (1.01, 1.39)	1.22 ± 0.46 (1.02, 1.41)	0.17
Ventilation defect percent (%)				
Normal (N=16)	Right	4.21 ± 6.24 (0.89, 7.54)	4.15 ± 6.07 (0.91, 7.38)	0.67
	Left	3.97 ± 2.71 (2.53, 5.42)	3.95 ± 2.36 (2.69, 5.21)	0.53
Asthma (N=5)	Right	18.32 ± 12.30 (3.05, 33.59)	19.34 ± 10.94 (5.75, 32.93)	0.31
	Left	22.37 ± 14.36 (4.54, 40.20)	23.46 ± 13.65 (6.51, 40.41)	1.00
Cystic fibrosis (N=24)	Right	22.03 ± 15.19 (15.62, 28.45)	22.31 ± 15.61 (15.72, 28.90)	0.80
	Left	19.90 ± 13.57 (14.17, 25.62)	20.16 ± 13.71 (14.37, 25.95)	0.82

Data are presented as mean ± standard deviation (95% Confidence levels). The volumetric and functional measurements extracted from the deep learning predicted masks and the reference masks were compared in each subject group using two-sided Wilcoxon signed-rank test. A $p < 0.05$ is considered statistically significant.