Generating high-resolution synthetic CT from lung MRI with ultrashort echo-times: initial evaluation in cystic fibrosis

Arthur Longuefosse, Julien Raoult, Ilyes Benlala, Baudouin Denis de Senneville, Thomas Benkert, Julie Macey, Stéphanie Bui, Patrick Berger, Gilbert Ferretti, Jean-Yves Gaubert, et al.

Generating high-resolution synthetic CT from lung MRI with ultrashort echo-times: initial evaluation in cystic fibrosis

Arthur Longuefosse, Julien Raoult, Ilyes Benlala, Baudouin Denis de Senneville, Thomas Benkert, Julie Macey, Stéphanie Bui, Patrick Berger, Gilbert Ferretti, Jean-Yves Gaubert, Renan Liberge, Antoine Hutt, Baptiste Morel, François Laurent, Fabien Baldacci, Gaël Dournes

Summary statement

Synthetic CTs generated by AI from lung MRI with ultrashort echo times had better image quality than the original MR images and showed strong agreement with real CTs for the depiction of Cystic Fibrosis-related pulmonary alterations.

Key Results:

1. In this retrospective study, synthetic CT images produced by Generative Adversarial Networks using lung MRI with ultrashort echo-times (UTE-MRI) from patients with Cystic Fibrosis (CF) had a higher contrast-to-noise ratio than MR images, a lower noise level than CT, and the lowest level of artifacts among all 3 image types ($P \leq .001$).

2. Almost perfect concordance was observed between real and synthetic CT images for the assessment of CF-related pulmonary structural alterations (ICC, ≥ 0.92).

Abbreviations:

CF, cystic fibrosis

UTE-MRI, ultrashort echo time MRI

PFT, pulmonary function test

GAN, generalized adversarial networks

ICC, intraclass correlation coefficient

Abstract

Background: Lung MRI with ultrashort echo times (UTE-MRI) enables high-resolution and radiation-free morphological imaging, however, its imaging quality is still lower than CT.

Purpose: To assess the image quality and clinical applicability of synthetic CTs generated from lung UTE-MRI using Generative Adversarial Networks (GAN).

Methods and Materials: This retrospective study included patients with Cystic Fibrosis (CF) who underwent both UTE-MRI and CT on the same day at six institutions between January 2018 and December 2022. The 2D GAN algorithm was trained using paired MR-to-CT slices and tested, along with an external dataset. Imaging quality was assessed quantitatively by measuring apparent contrast-to-noise-ratio, apparent signal-to-noise-ratio, and overall noise and qualitatively using visual scores for features including artifacts. Two readers evaluated CF-related structural abnormalities and used them to determine clinical Bhalla scores.

Results: The training, test, and external datasets were comprised of 82 CF patients (mean age, 21 ± 11 [SD]; 42 males), 28 patients (mean age, 18 ± 11; 16 males), and 46 patients (mean age, 20 ± 11; 24 males), respectively. In the test dataset, the contrast-to-noise ratio of synthetic CTs (median, 303 [interquartile range: 221, 382]) was higher than that of UTE-MRI (9.3 [6.6, 35]; $P < .001$). The median signal-to-noise ratio was similar between the synthetic and corresponding CT (88 [84, 92] vs 88 [86, 91]; $P = 0.96$). Synthetic CT had a lower noise level than CT (median score, 22 [18, 30] vs 42 [32, 50]; $P < .001$) and the lowest level of artifacts (median score, 0 [0, 0]; $P \leq .001$). The concordance between Bhalla scores for synthetic and real CTs was almost perfect (ICC, ≥ 0.92).

Conclusion: Synthetic CTs showed almost perfect concordance with CTs for the depiction of Cystic Fibrosis-related pulmonary alterations and had better image quality than lung MRI with ultrashort echo-times.

INTRODUCTION

CT is the standard of reference for morphological imaging of the lung, owing to its excellent spatial resolution and contrast between air and intrapulmonary structures. However, CT imaging is based on ionizing radiations, which may raise concerns in chronic conditions when repeated imaging is needed (1). Alternatively, lung MRI is a radiation-free imaging modality, but it remains a technical challenge owing to such things as the low proton density of this tissue, magnetic susceptibility artifacts, and respiratory motions. Lung MRI with ultrashort echo-times (UTE-MRI) (2,3) represents a recent breakthrough which helps with catching up the very fast decay of the lung MR signal and reducing susceptibility artefacts. As UTE-MRI does not require a contrast material injection or inhalation, it could be used to evaluate Cystic Fibrosis (CF) in the lungs of children, a population for which a radiation-free examination would be desirable. Indeed, previous studies have consistently demonstrated a better depiction of structural alterations with UTE-MRI than with conventional T1-weighted MR sequences (4,5). Furthermore, several international position papers have recently agreed on the clinical use of proton MRI in the longitudinal follow-up of patients with CF, with the introduction of UTE-MRI (1,6,7). However, UTE-MRI imaging quality remains inferior to that of CT. It displays a different imaging texture than the standard of reference CT, has a lower contrast than CT (5) and features specific artifacts such as blurring or motions (8), questioning its clinical adoption without adept readers(9–11). Indeed, there is a need for training to decipher CT semantic from UTE-MRI with confidence.

Nevertheless, artificial intelligence (AI) is currently changing the landscape of imaging tools available. Recently, deep learning with generative adversarial networks (GAN) have been developed and provides a framework for estimating generative models. Amongst the potential applications of GAN (12), there is a possibility of MR-to-CT translations. Lung MRI signal intensity conversion into CT-like attenuation values have been reported, demonstrating an interest in radiotherapy planning with PET-MRI (13,14). However, the feasibility of GAN to get high-resolution morphological imaging from lung UTE-MRI, that would reach CT-like quality, has not been evaluated yet. Moreover, the clinical impact of GAN for improving MRI-based evaluations of structural alterations in CF is unknown.

The aim of this study was to generate synthetic CTs from UTE using GAN and compare the image quality and clinical applicability of these synthetic images to their corresponding UTE-MRI and real CT scans.

MATERIAL AND METHODS

Study Sample

This feasibility AI study was retrospective and involved two different datasets(15), collected between January 2018 and December 2022.

For model development and test, data were retrieved from medical records at the University Hospital of Bordeaux. Inclusion criteria were: 1) an age ≥ 8 years, 2) a diagnosis of CF determined by genetic testing and/or sweat chloride tests, 3) availability of non-contrast-enhanced CT, MRI and PFT performed on the same day during an annual examination(16). Exclusion criteria were any contraindication to complete lung MRI. All patients and/or their parents provided written informed consent according to the local Ethics committee (CHUBX2020RE0267). The data were randomly split 75%/25% using Research Randomizer® into a training and test datasets. As a first-attempt study, an arbitrary minimum of 25 patients was calculated to assess a concordance in CF scorings > 0.6, with a risk alpha of 0.05 and a power of 90%.

For external validation(15), 13 geographically distinct Institutions were screened amongst a multicenter trial (*ClinicalTrials.gov*: NCT03357562). Five of them conferred their agreement for an ancillary analysis of their prospectively collected data (Figure 1). Inclusion and exclusion criteria were the same as those for the training and test datasets, however, PFT data were not collected.

Imaging Protocols

MRI was performed with a 1.5 Tesla unit (MAGNETOM Aera, Siemens Healthcare), except for one center where MRI was done on a 3T scanner (MAGNETOM Skyra, Siemens Healthcare). In all centers, the UTE-MRI Spiral Volume Interpolated Breathold Examination (VIBE) sequence was implemented for evaluation(17) (Supplemental Method E1). The basic sequence parameters were TR/TE/flip angle=4.1ms/0.07ms/5°; slice thickness=1mm; pixel size=1x1mm; acquisition time from 6 to 8 minutes with a prospective respiratory gating at end normal expiration. No contrast agent or sedation was used (Supplemental Method E1).

CT scans were done on 4 machine brands from two main manufacturers, *i.e.* GE and Siemens Healthineers. In all centers, patients were instructed to hold their breath at functional residual

capacity (end normal expiration)(17,18). CTDIvol ranged from 0.6 to 7.5 mGy and Dose Length Product from 13 to 260 mGy.cm (Supplemental Table E1).

Generation of Synthetic CT using GAN

The general GAN principle is described and illustrated in Figure 2. 2D GAN architecture was set up with Semantic Image Synthesis with Spatially-Adaptive Normalization (SPADE) (https://github.com/alexandonian/contrastive-feature-loss)(19). Details of the framework to generate synthetic CT from UTE-MRI using the Training dataset (Supplemental Tables E1-E2) are provided in Supplemental Method E2(19–23).

Evaluation of image quality

All examinations from the Test and External datasets were anonymized and randomly assigned for evaluation. For quality assessment, a reader with 4 years (J.R.) and 10 years (G.D.) of experience in chest imaging made the evaluations according to previously published methods(3,24–26).Quantitative assessment of apparent contrast-to-noise ratio, signal-to-noise ratio(2,3) and noise(24) were measured. Qualitative analysis was done(25,26) by assessing a total visibility score (range 0-30); total sharpness score (range 0-24); and total artefact score (range 0-45) (Supplemental Method E3-E4; Supplemental Tables E3).

Evaluation of CF structural alterations at the segmental level

At the segmental level, the presence of structural alterations was assessed using a binary scale (0, absent; 1, present) by G.D. (Supplemental Methods E5), i.e. peribronchial thickening, bronchiectasis, mucus plugs, bronchiolar impactions, collapse/consolidation, emphysema, bulla, and sacculation/abscess(27).

Concordance and reproducibility to assess the Bhalla score.

The clinical validity of synthetic CT was assessed using the original version of the Bhalla score(28). The scoring system represents the severity of CF structural alterations and ranges from 0 to 25 (Supplemental Method E6, Supplemental Table E4). The junior reader (J.R.) had used the Bhalla score with CT but not MRI, while the senior reader (G.D.) had experience

using the Bhalla score in both CT and MRI(17). Bhalla scores were determined independently, twice, by each reader for UTE-MRI, Synthetic CT, and CT. The Bhalla scores from the senior's CT first reading session were used as the reference standard.

Statistical analysis

Statistical analysis was carried out by P.B. (20 years of experience with statistical analysis) using MedCalc Software 20.218 Ltd. Normality of data distribution was tested with the Shapiro Wilk test. Quantitative data were expressed as means ± standard deviation or median with [interquartile range] according to the normality or non-normality of distributions, respectively. Categorical data as number of patients and percentages. Comparisons of means was done with Student's t test. Comparisons of medians were done with Mann-Whitney test and paired medians with the Friedman test with post-hoc Dunn's test. For categorical variables, percentages were compared using the Chi-Square test. Agreement of structural alteration depiction and concordance in Bhalla scores were assessed using the weighted-kappa coefficient (κ_w) and intraclass correlation coefficient (ICC), respectively. Kappa and ICC values was scored as slight (≤0.20), fair (≥0.20 and <0.40), moderate ((≥0.40 and <0.60), substantial (≥0.60 and <0.80), and almost perfect (≥ 0.81)(29,30). The relationship between Bhalla scores and PFT were assessed using Pearson correlation.

RESULTS

Study Population

A total of 110 consecutive CF patients completed both lung CT and UTE-MRI during their annual follow-up. All patients met the criteria of this study and thus none were excluded (Figure 1). Of these 110 patients, 82 (75%) were randomly assigned to the training dataset (mean age, 21 years ± 11 [SD]; 42 males, 40 females) and a total of 33 002 paired MR and CT slices were used to train the GAN algorithm (Supplemental Table E2). The test cohort included 28/110 (25%) CF patients of which 16 were male and 12 were female. The mean age in the test dataset was 18 years ± 11 with an age range of 8 to 50 years old. The mean Bhalla score for the test dataset was 16.6 ± 5.3. Additional patient characteristics are presented in Table 1.

From the 5 external Institutions who participated in this AI study, 46 CF patients who completed both lung CT and UTE-MRI on the same day were included in our external dataset. The mean age of patients in the external dataset was 20 years ± 11 and the mean CT Bhalla score was 12.6 ± 2.5 (Table 1).

Evaluation of image quality

Quantitative assessment of image quality.

In the test cohort, no difference was observed in the Signal-to-noise ratio between all imaging modalities ($P = 0.63$). Conversely, the median Contrast-to-noise ratio for synthetic CT images was 303 [221, 382] compared to a median of 9.3 [6.6, 35] for UTE-MRI ($P < .001$). However, the overall noise was higher in real CT compared to MRI (median, 42 [32, 50] vs 26 [22, 30], respectively; $P < .001$) (Table 2, Supplemental Table E5, Figure 3).

Qualitative assessment of image quality

UTE-MRI had lower qualitative scores in regard to the blurring of contours (median score, 10.5 [9, 12]) compared to real CT (median, 22 18, 24]) or synthetic CT (median score, 22 [18, 24]; $p < .001$)). Synthetic CT had the lowest total artifact score (median score, 0 [0, 0]; $P < .001$) (Table 2). Beam hardening artifacts were noticed only on CT, in 17/28 (60%) patients. Only UTE-MRI displayed slight-to-moderate motion artifacts in 14/28 (50%) patients with imperfect respiratory synchronization (Supplemental Table E6, Figure 4),

Confirmatory results in the external dataset

In the External dataset, the median Contrast-to-noise ratio for synthetic CT images was 364 [258, 448], without significant difference with Test dataset ($p = 0.22$). Using MRI, a slight aliasing that did not overlap with the lung area was noticed in 1 female (Figure 4). As compared to the Test dataset, no significant difference in External dataset's imaging quality results was noticed, such as total sharpness score (median score, 22 [20, 24]; $P = 0.62$) or total artefact score (median score, 0 [0, 0]; $P = 0.43$) (Supplemental Tables E7-E8).

Evaluation of CF structural alterations at the segmental level

There were 504 lung segments assessed among the 28 CF patients in the test dataset. A better agreement was observed between CT and Synthetic CT than between CT and UTE for the structural alterations bronchiectasis (CT vs synthetic CT $\kappa = 0.91$, CT vs UTE-MRI $\kappa = 0.70$), wall thickening (CT vs synthetic CT $\kappa = 0.89$, CT vs UTE-MRI $\kappa = 0.70$), and mucus plugs (CT vs synthetic CT $\kappa = 0.98$, CT vs UTE-MRI $\kappa = 0.91$), without overlap in 95% confidence intervals (Table 3). Additionally, fewer false negative structural alterations were observed on Synthetic CT (Supplemental Table E8). For instance, 55/252 (21%) lung segments with bronchiectasis were missed on UTE-MRI in the test cohort, although 16/252 (6%) lung segments with were missed on Synthetic CT. Moreover, fewer false positive structural alterations were observed on synthetic CT (Supplemental Table E8). For instance, on UTE-MRI there were 12/132 (9%) lung segments with a false positive depiction of mucus plugs compared to 1/132(0.7%) lung segment with a false positive on synthetic CT.

There were 828 lung segments assessed among the 46 CF patients in the external dataset, and agreement regarding structural alterations was better between CT and Synthetic CT compared to CT and UTE for instance regarding bronchiectasis ($\kappa = 0.97$ and $\kappa = 0.79$, respectively) (Table 3). Furthermore, a reduction in both false positive and false negative depictions of structural alterations such as severe (Figure 4) or mild (Figure 5) bronchiectasis was observed on Synthetic CT (Table 3, Supplemental Table E9).

Bhalla score concordance between readers and correlation with pulmonary function tests.

For the junior reader, the UTE-MRI Bhalla score showed substantial concordance with the real CT reference Bhalla score (ICC, 0.70) in the test dataset. Alternatively, the concordance of the Synthetic CT or real CT Bhalla scores with the reference CT Bhalla score were excellent (ICC, 0.92 and ICC, 0.93, respectively).

In the external dataset, substantial concordance was observed between UTE Bhalla scores assigned by the junior reader and the real CT reference Bhalla score (ICC, 0.77), and an almost perfect concordance between Synthetic CT and real CT reference Bhalla scores (ICC, 0.93) (Table 4, Figure 5). For the senior reader, there was excellent concordance between both

UTE (ICC, 0.95) and Synthetic CT (ICC, 0.97) Bhalla scores and the real CT reference Bhalla score.

Synthetic CT, and UTE-MRI Bhalla scores assessed by the senior reader demonstrated positive correlations with Forced Expiratory Volume in 1 second percentage predict (FEV1%p) (r = 0.62, $P < .001$; r = 0.62, $P < .001$; r = 0.60, $P < .001$; respectively) and FEV1/Forced Vital Capacity (FVC) (r = 0.52, $P = .005$; r = 0.53, $P = .004$; r = 0.45, $P = .01$; respectively). However, UTE Bhalla scores determined by the junior reader alone did not correlate with the obstructive pattern, assessed by the FEV1/FVC ratio (r = 0.33; $P = .08$) (Supplemental Table E10).

The intra-observer reproducibility of UTE, Synthetic CT or CT evaluations was almost perfect for the junior reader (ICC ≥ 0.90) and the senior reader (ICC ≥ 0.98) (Supplemental Table E11).

DISCUSSION

Lung MRI with ultrashort echo times (UTE-MRI) is a recent breakthrough for morphological MR imaging. However, its imaging quality is still inferior to CT. The aim of the study was to improve UTE-MRI quality by using post-processing with Generative Adversarial Networks (GAN). The results showed that the median contrast-to-noise ratio for synthetic CT was 303 [221, 382], higher than that of UTE-MRI (median, 9.3 [6.6, 35]; $P < .001$). Compared to the other images, synthetic CT had the lowest level of artifacts (median score, 0 [0, 0] $P \leq .001$), The concordance between Bhalla scores for synthetic and real CTs was almost perfect for both a senior reader (ICC, 0.99) and a junior reader without any experience in UTE-MRI scoring (ICC, 0. 92). Finally, Synthetic CT Bhalla scores correlated with pulmonary function tests such as forced expiratory volume in 1 second percentage predicted (r, 0.62; $p < .001$).

To the best of our knowledge, this study is the first to report the use of GAN for generating high-resolution CT-like morphological imaging from lung UTE-MRI. Previous lung MRI studies with GAN have focused on converting the MRI signal into CT-like attenuation values to enable correcting the metabolic information of PET-MRI from low resolution images (13,14). Indeed, those evaluations were not done with UTE-MRI, and the resolution was of 3

mm slice thickness at best. Thus, an evaluation of morphological information was not reported.

Herein, GAN was able to generate 3D isotropic synthetic lung CT with a millimetric resolution. Also, we found that heterogeneously distributed MRI artifacts, such as slight-to-moderate motion artifacts, were removed after GAN processing. Of note, AI appeared able to correct them without affecting the global imaging features. Conversely, we found that some structural alterations that were visually missed by the expert reader on UTE alone were depicted on Synthetic CT. Those features were consistently confirmed as true positives on corresponding CT. Generation of synthetic CT images by GAN did not introduce additional false positive findings in both our test and external dataset.

In this study, CT scans protocols were not uniform, questioning the results whether only CT at high doses may have been used. Nevertheless, the study was retrospective, and the reduction of CT doses has been shown to not alter the CT morphological information in CF (31). In addition, the resulting GAN model was not dependent on a single manufacturer or kernel, and the integrated GAN model showed good generalizability, with similar imaging quality results between a Test and External datasets ($P \geq 0.12$).

In addition, the clinical robustness and applicability of the method were evaluated through the assessment of the Bhalla score. The concordance of the Bhalla score assigned to Synthetic CT and real CT was almost perfect according to both an expert (ICC, 0.99) and a non-expert reader (ICC, 0.92). Of note, the non-expert reader had a lower concordance by using UTE alone (ICC, 0.70). These findings were constant across a Test dataset and an external dataset of CF patients' examinations from geographically distinct Institutions (15). Also, these ICC values were comparable to those reported in the literature with CT, showing ICC of 0.96 [0.92, 0.98] to assess the Bhalla score using CT (31).

Therefore, synthetic CT appeared to enhance the applicability of lung MRI in the clinical setting. It may open the door to a new era of lung imaging without continuing dependence upon ionizing radiation, which may be applicable to diseases beyond CF, and perhaps beyond the chest. Also, these initial results may pave the way for more standardization of MRI examinations and translating CT methods of evaluation to MRI (7,32). One could discuss that CT is also being revolutionized by emerging evolutions for reducing radiation dose while maintaining image quality (33). Nevertheless, MRI has no ionizing radiation. Also, synthetic CT could enhance MRI as a "one-stop-shop" ionizing radiation-free examination, with

possibility to combine high resolution morphology to functional information (34,35) in a reasonable acquisition time of less than 15 minutes (6). Next studies may address the relative contribution of these techniques in the future.

The study had several limitations. First, the study was retrospective, and a larger prospective study involving several junior and senior readers should be done to assess the clinical robustness further. Second, a single scoring system was evaluated, and other scoring methods may be worth evaluating for larger applications. Third, the technique was based on 2D algorithms, and there may be ways for improvements such as 3D algorithms. Fourth, MRI is inferior to CT in terms of cost and acquisition time. Fifth, the multicenter study was done by using a single MR sequence scheme and the generalizability to other UTE or ZTE sequence schemes remains to be explored.

CONCLUSION

Synthetic CTs showed almost perfect concordance with CTs for the depiction of CF-related pulmonary alterations and had better image quality than lung MRI with ultrashort echo-times.

Acknowledgement.

This study was conducted in the framework of the University of Bordeaux's IdEx "Investments for the Future" program RRI "IMPACT" which received financial support from the French government.

The authors would like to thank Mrs Rkia Achkir and Dr Solenn Toupin for their technical support.

References.

1. Hatabu H, Ohno Y, Gefter WB, et al. Expanding Applications of Pulmonary MRI in the Clinical Evaluation of Lung Disorders: Fleischner Society Position Paper. Radiology 2020;297(2):286–301. doi: 10.1148/radiol.2020201138.

2. Johnson KM, Fain SB, Schiebler ML, Nagle S. Optimized 3D ultrashort echo time pulmonary MRI. Magn Reson Med 2013;70(5):1241–1250. doi: 10.1002/mrm.24570.

3. Dournes G, Grodzki D, Macey J, et al. Quiet Submillimeter MR Imaging of the Lung Is Feasible with a PETRA Sequence at 1.5 T. Radiology 2015;276(1):258–265. doi: 10.1148/radiol.15141655.

4. Dournes 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(11):3811–3820. doi: 10.1007/s00330-016-4218-5.

5. Renz DM, Herrmann K-H, Kraemer M, et al. Ultrashort echo time MRI of the lung in children and adolescents: comparison with non-enhanced computed tomography and standard post-contrast T1w MRI sequences. Eur Radiol 2022;32(3):1833–1842. doi: 10.1007/s00330-021-08236-7.

6. Dournes G, Walkup LL, Benlala I, et al. The Clinical Use of Lung MRI in Cystic Fibrosis: What, Now, How? Chest 2021;159(6):2205–2217. doi: 10.1016/j.chest.2020.12.008.

7. Ciet P, Bertolo S, Ros M, et al. State-of-the-art review of lung imaging in cystic fibrosis with recommendations for pulmonologists and radiologists from the "iMAging managEment of cySTic fibROsis" (MAESTRO) consortium. Eur Respir Rev 2022;31(163):210173. doi: 10.1183/16000617.0173-2021.

8. Papp D, Elders B, Wielopolski PA, et al. Lung parenchyma and structure visualisation in paediatric chest MRI: a comparison of different short and ultra-short echo time protocols. Clin Radiol 2023;78(4):e319–e327. doi: 10.1016/j.crad.2022.12.020.

9. Gräfe D, Anders R, Prenzel F, et al. Pediatric MR lung imaging with 3D ultrashort-TE in free breathing: Are we past the conventional T2 sequence? Pediatr Pulmonol 2021;56(12):3899–3907. doi: 10.1002/ppul.25664.

10. Gross JE, McCown MY, Okorie CUA, et al. ATS Core Curriculum 2021. Pediatric Pulmonary Medicine: Pulmonary Infections. ATS Sch. 2021;2(3):452–467. doi: 10.34197/ats-scholar.2021-0034RE.

11. Tiddens HAWM, Kuo W, van Straten M, Ciet P. Paediatric lung imaging: the times they are a-changin'. Eur Respir Rev 2018;27(147):170097. doi: 10.1183/16000617.0097-2017.

12. Yi X, Walia E, Babyn P. Generative adversarial network in medical imaging: A review. Med Image Anal 2019;58:101552. doi: 10.1016/j.media.2019.101552.

13. Lenkowicz J, Votta C, Nardini M, et al. A deep learning approach to generate synthetic CT in low field MR-guided radiotherapy for lung cases. Radiother Oncol 2022;S0167-8140(22)04260-8. doi: 10.1016/j.radonc.2022.08.028.

14. Baydoun A, Xu K, Yang H, et al. Dixon-based thorax synthetic CT generation using Generative Adversarial Network. Intell Based Med 2020;3–4:100010. doi: 10.1016/j.ibmed.2020.100010.

15. Park SH, Han K. Methodologic Guide for Evaluating Clinical Performance and Effect of Artificial Intelligence Technology for Medical Diagnosis and Prediction. Radiology 2018;286(3):800–809. doi: 10.1148/radiol.2017171920.

16. Mucoviscidose protocole national de diagnostic et de soins pour une maladie rare. Haute Autorité de Santé. https://www.has-sante.fr/upload/docs/application/pdf/2017-09/pnds_2017_vf1.pdf. Published July 17, 2017. Accessed March 1, 2023.

17. Dournes G, Yazbek J, Benhassen W, et al. 3D ultrashort echo time MRI of the lung using stack-of-spirals and spherical k-Space coverages: Evaluation in healthy volunteers and parenchymal diseases. J Magn Reson Imaging. 2018;48(6):1489–1497. doi: 10.1002/jmri.26212.

18. Loeve M, Lequin MH, de Bruijne M, et al. Cystic fibrosis: are volumetric ultra-low-dose expiratory CT scans sufficient for monitoring related lung disease? Radiology 2009;253(1):223–229. doi: 10.1148/radiol.2532090306.

19. Park T, Liu M-Y, Wang T-C, Zhu J-Y. Semantic Image Synthesis with Spatially-Adaptive Normalization. arXiv; 2019. http://arxiv.org/abs/1903.07291. Published November 5, 2019. Last accessed December 26, 2022.

20. Denis de Senneville B, Zachiu C, Ries M, Moonen C. EVolution: an edge-based variational method for non-rigid multi-modal image registration. Phys Med Biol 2016;61(20):7377–7396. doi: 10.1088/0031-9155/61/20/7377.

21. Hofmanninger J, Prayer F, Pan J, Röhrich S, Prosch H, Langs G. Automatic lung segmentation in routine imaging is primarily a data diversity problem, not a methodology problem. Eur Radiol Exp 2020;4(1):50. doi: 10.1186/s41747-020-00173-2.

22. Andonian A, Park T, Russell B, Isola P, Zhu J-Y, Zhang R. Contrastive Feature Loss for Image Prediction. arXiv; 2021; doi: 10.48550/ARXIV.2111.06934. Published November 12, 2021. Last accessed December 30, 2022.

23. Kingma DP, Ba J. Adam: A Method for Stochastic Optimization. arXiv; 2014; doi: 10.48550/ARXIV.1412.6980. Published December 22, 2014. Last accessed December 30, 2022.

24. Symons R, Pourmorteza A, Sandfort V, et al. Feasibility of Dose-reduced Chest CT with Photon-counting Detectors: Initial Results in Humans. Radiology 2017;285(3):980–989. doi: 10.1148/radiol.2017162587.

25. Ohno Y, Koyama H, Yoshikawa T, et al. Pulmonary high-resolution ultrashort TE MR imaging: Comparison with thin-section standard- and low-dose computed tomography for the assessment of pulmonary parenchyma diseases. J Magn Reson Imaging 2016;43(2):512–532. doi: 10.1002/jmri.25008.

26. Darçot E, Delacoste J, Dunet V, et al. Lung MRI assessment with high-frequency noninvasive ventilation at 3 T. Magn Reson Imaging 2020;74:64–73. doi: 10.1016/j.mri.2020.09.006.

27. Hansell DM, Bankier AA, MacMahon H, McLoud TC, Müller NL, Remy J. Fleischner Society: glossary of terms for thoracic imaging. Radiology 2008;246(3):697–722. doi: 10.1148/radiol.2462070712.

28. Bhalla M, Turcios N, Aponte V, et al. Cystic fibrosis: scoring system with thin-section CT. Radiology 1991;179(3):783–788. doi: 10.1148/radiology.179.3.2027992.

29. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33(1):159–174.

30. Fleiss JL, Cohen J. The Equivalence of Weighted Kappa and the Intraclass Correlation Coefficient as Measures of Reliability. Educational and Psychological Measurement. 1973;33(3):613–619. doi: 10.1177/001316447303300309.

31. Ernst CW, Basten IA, Ilsen B, et al. Pulmonary disease in cystic fibrosis: assessment with chest CT at chest radiography dose levels. Radiology 2014;273(2):597–605. doi: 10.1148/radiol.14132201.

32. Dournes G, Hall CS, Willmering MM, et al. Artificial intelligence in computed tomography for quantifying lung changes in the era of CFTR modulators. Eur Respir J 2022;59(3):2100844. doi: 10.1183/13993003.00844-2021.

33. Horst KK, Hull NC, Thacker PG, et al. Pilot study to determine whether reduced-dose photon-counting detector chest computed tomography can reliably display Brody II score imaging findings for children with cystic fibrosis at radiation doses that approximate radiographs. Pediatr Radiol 2023; doi: 10.1007/s00247-022-05574-6.

34. Wielpütz MO, Puderbach M, Kopp-Schneider A, et al. Magnetic resonance imaging detects changes in structure and perfusion, and response to therapy in early cystic fibrosis lung disease. Am J Respir Crit Care Med 2014;189(8):956–965. doi: 10.1164/rccm.201309-1659OC.

35. Veldhoen S, Heidenreich JF, Metz C, et al. Three-dimensional Ultrashort Echotime Magnetic Resonance Imaging for Combined Morphologic and Ventilation Imaging in Pediatric Patients With Pulmonary Disease. J Thorac Imaging 2021;36(1):43–51. doi: 10.1097/RTI.0000000000000537.

TABLES

Table 1. Patient characteristics.

	Test dataset (n=28)	**External dataset (n=46)**	**p-value**
Age (years)	18 ± 11	20 ± 11	0.45
Sex			0.67
Male	16 (57%)	24 (52%)	
Female	12 (43%)	22 (48%)	
BMI (kg.m^{-2})	19 ± 3	20 ± 2.6	0.13
DeltaF508 mutation			0.37
Homozygous	15 (53%)	28 (61%)	
Heterozygous	13 (47%)	16 (35%)	
Other mutation	0 (0%)	2 (4%)	
Pulmonary function test			
FEV1 (%p)	82 ± 22	NA	NA
FEV1/FVC	78 ± 14	NA	NA
RV (%p)	131 ± 46	NA	NA
TLC (%p)	97 ± 15	NA	NA
CT Bhalla score	16.6 ± 5.3	12.6 ± 2.5	<0.001

Note.—Data are mean ± standard deviation for continuous data and number of patients with percentages in parentheses for categorical data. BMI = body mass index, FEV1 = forced expiratory volume in 1 second, FVC = forced vital capacity, RV = residual volume, TLC = total lung capacity, NA = not applicable.

Table 2. Quantitative and qualitative assessment of image quality.

Test dataset (n=28 patients)	**UTE-MRI**		**Synthetic CT**		**CT**		**p-value**
Quantitative analysis							
Contrast-to-noise ratio	**9.3****,†	[6.6, 35]	303	[221, 382]	247	[210, 323]	**<0.001**
Signal-to-noise-ratio	88	[87, 91]	88	[84, 92]	88	[86, 91]	0.63
Overall Noise	22	[18, 31]	26	[22, 30]	**42***,†	[32, 50]	**<0.001**
Qualitative analysis							
Total visibility score	**27****,†	[24, 29]	29	[27, 30]	29	[27, 30]	**<0.001**
Total sharpness score	**10.5****,†	[9, 12]	22	[18, 24]	22	[18, 24]	**<0.001**
Total artifact score	**0.5****	[0, 1]	0	[0, 0]	**1***	[0, 1]	**<0.001**
External dataset (n=46 patients)	**UTE-MRI**		**Synthetic CT**		**CT**		**p-value**
Quantitative analysis							
Contrast-to-noise ratio	**13****,†	[7, 20]	364	[258, 448]	245	[155, 324]	**<0.001**
Signal-to-noise ratio	90	[88, 93]	90	[88, 92]	90	[87, 92]	0.75
Overall Noise	21.5	[14, 45]	23	[18, 28]	**39***,†	[20, 46]	**0.002**
Qualitative analysis							
Total visibility score	**28****,†	[26, 29]	29	[28, 30]	29	[28, 30]	**<0.001**
Total sharpness score	**11****,†	[9, 18]	22	[20, 24]	22	[18, 24]	**<0.001**
Total artifacts score	**1****	[0, 3]	0	[0, 0]	**0.5***	[0, 1]	**<0.001**

Note: The (minimum to maximum) range of visibility, sharpness, and artifact total scores are (0 to 30), (0 to 24) and (0 to 45), respectively.

For quantitative and qualitative analysis, data are medians with [interquartile range]. Overall comparison of medians was assessed with Friedman test and post hoc Dunn's test. * indicates a p-value ≤ .001 between CT and Synthetic CT, ** indicates a p-value ≤ .001 between UTE-MRI and Synthetic CT and † indicates a p-value ≤ .001 between UTE-MRI and CT.

UTE-MRI = lung MRI with ultrashort echo times

Table 3. Strength of Agreement between reader evaluations of structural alterations at a segmental level.

Test dataset (n=504 segments in 28 patients)		**CT vs UTE-MRI**		**CT vs Synthetic CT**	
		κ_w	**95% CI**	**κ_w**	**95% CI**
Peribronchial thickening					
Present	277 (54.9%)	0.70	[0.64, 0.77]	0.89	[0.85,0.93]
Bronchiectasis					
Present	252 (50%)	0.76	[0.71, 0.82]	0.91	[0.88, 0.95]
Central mucus					
Present	131 (25.9%)	0.91	[0.87, 0.95]	0.98	[0.97,1.00]
Peripheral mucus					
Present	179 (35.5%)	0.90	[0.86, 0.94]	0.98	[0.96, 0.99]
Collapse/Consolidation					
Present	12 (2.3%)	0.90	[0.77, 1.00]	>0.99	NA
Emphysema					
Present	0 (0%)	NA	NA	NA	NA
Sacculation/Abscess					
Present	3 (0.5%)	>0.99	NA	>0.99	NA
Bulla					
Present	0 (0%)	NA	NA	NA	NA
External dataset (n=828 segments in 46 patients)		**CT vs UTE-MRI**		**CT vs Synthetic CT**	
		κ_w	**95% CI**	**κ_w**	**95% CI**
Peribronchial thickening					
Present	703 (84.9%)	0.81	[0.76, 0.87]	0.94	[0.91, 0.97]
Bronchiectasis					
Present	677 (81.2%)	0.79	[0.74, 0.84]	0.97	[0.95, 0.99]
Central mucus					
Present	491 (59.2%)	0.86	[0.82, 0.89]	0.94	[0.91, 0.96]
Peripheral mucus					

Present	515 (62.1%)	0.83	[0.80, 0.87]	0.94	[0.92, 0.97]
Collapse/Consolidation					
Present	29 (3.5%)	0.94	[0.88, 1.00]	0.98	[0.94, 1.00]
Emphysema					
Present	0 (0%)	NA	NA	NA	NA
Sacculation/Abscess					
Present	9 (1%)	0.89	[0.75, 1.00]	>0.99	NA
Bulla					
Present	2 (0.2%)	0.66	[0.04; 1.00]	0.66	[0.04, 1.00]

Note: Evaluation of lung structural alterations according to the senior reader with CT as reference. Legends: UTE-MRI = lung MRI with ultrashort echo-times, CI = confidence interval, NA=not applicable.

Table 4. Assessment of Concordance between Bhalla scores assigned to each image type by a junior or senior reader.

	Test dataset (n=28 patients)			**External daataset (n=46 patients)**		
	Ref_CT vs. UTE-MRI	**Ref_CT vs. Synthetic CT**	**Ref_CT vs. CT**	**Ref_CT vs. UTE-MRI**	**Ref_CT vs. Synthetic CT**	**Ref_CT vs. CT**
Junior reader						
ICC	**0.70**	0.92	0.93	**0.77**	0.93	0.93
[95% CI]	[0.45, 0.84]	[0.83; 0.96]	[0.85; 0.97]	[0.60; 0.87]	[0.88; 0.96]	[0.88; 0.96]
Senior reader						
ICC	0.96	0.99	NA	0.95	0.97	NA
[95% CI]	[0.93; 0.98]	[0.99; 0.99]	NA	[0.92; 0.97]	[0.94; 0.98]	NA

Note: Ref_CT corresponds to the evaluation of lung structural alterations according to the senior reader on real computed tomography.

UTE-MRI = lung MRI with ultrashort echo-times, ICC = intraclass correlation coefficient, CI = confidence interval, NA = not applicable.

FIGURE LEGENDS

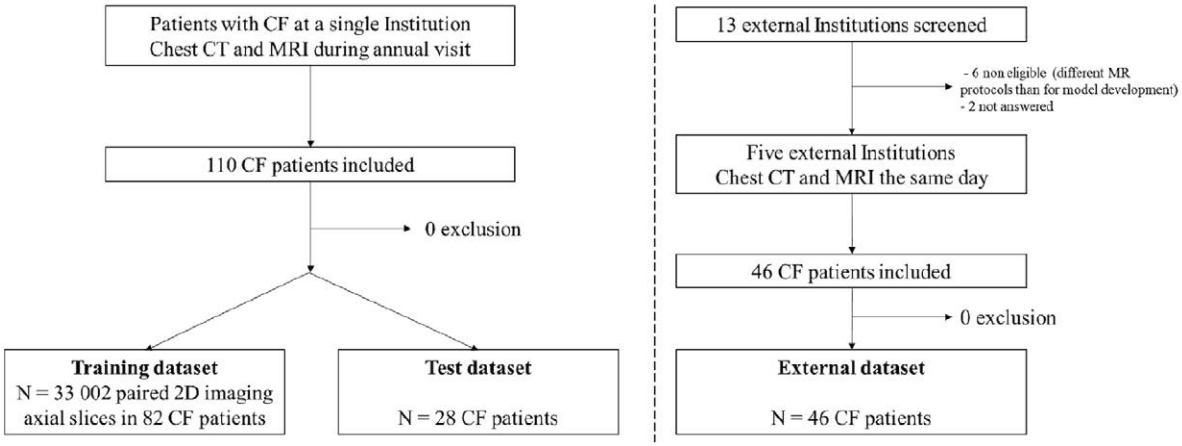

Figure 1. Study flow chart.

CF=cystic fibrosis, UTE-MRI = ultrashort echo-times.

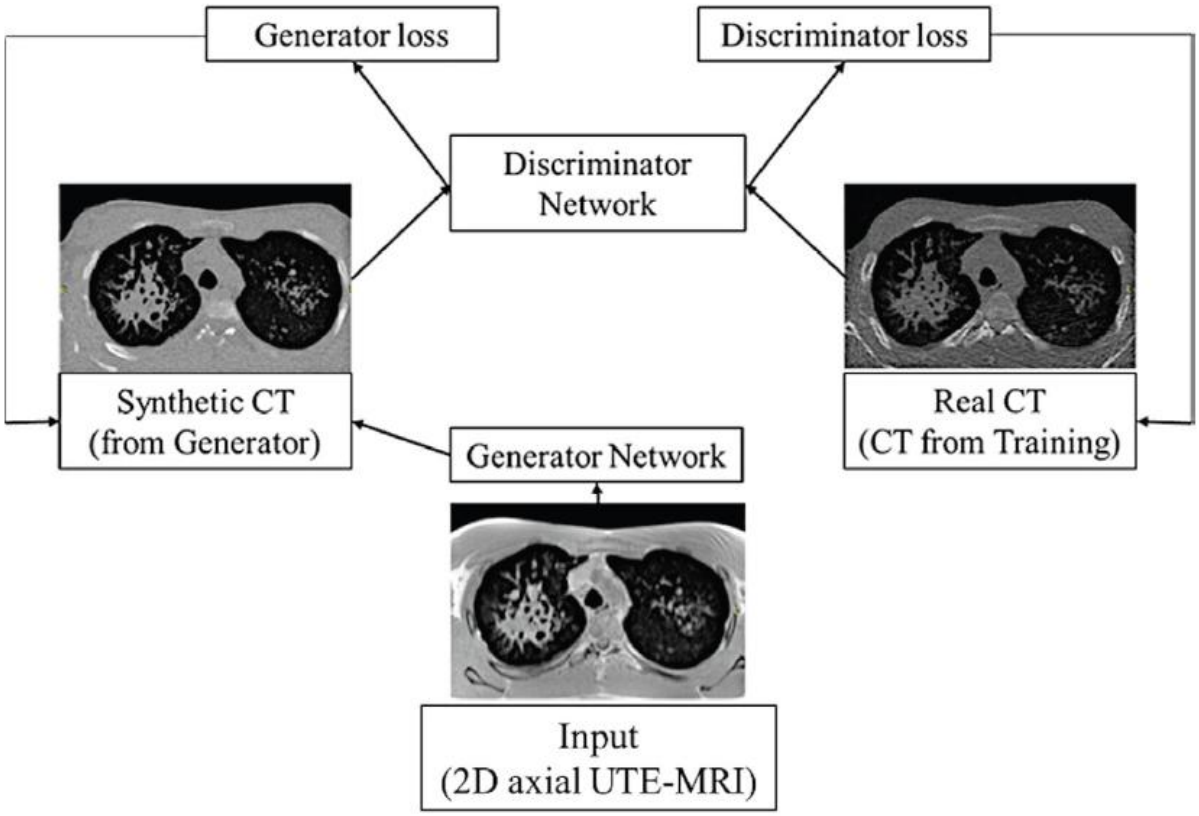

Figure 2. Generative Adversarial Network (GAN) principle.

A generator outputs synthetic samples given a variable input. Herein, the input corresponds to 2D axial slices from lung MRI with ultrashort echo-times (UTE-MRI) that belongs to the Training dataset.
A discriminator estimates the probability of a given sample to correspond to a real or a synthetic image. Herein, the real dataset refers to the corresponding 2D axial CT slice from the same patient that belongs to the Training dataset. The process reaches equilibrium when the discriminator can no longer distinguish the real CT slice from the synthetic CT slice.

Figure 3. Example images of the lung in a 12-year-old female with Cystic Fibrosis from the test dataset. The contrast-to-noise ratio was 8.5 on **(A)** Lung MRI with ultrashort echo-times (UTE-MRI), whereas it was 227 on **((B)** synthetic CT images generated by AI from UTE-MRI and 226 on (C) real CT examination with a clearer delineation of structural alterations, such as small bronchiolar impactions (white arrowhead) in the right lower lobe. the total sharpness score of the **(B)** synthetic CT was 22, better than **(A)** UTE-MRI for which it was equal to 14 and looks more like the **(C)** corresponding real CT scan for which it was equal to 22. On **(C)** low-dose CT image, note the presence of diffuse beam hardening artifacts due to the presence of the rib cage, some of them being indicated by white arrows.

Figure 4. Example images of the lung in a 21-year-old male with Cystic Fibrosis from the external validation dataset. On **(A)** lung MRI with ultrashort echo-times (UTE-MRI), there are motion artifacts on the right side (white arrows). These artifacts were not observed on **(B)** synthetic CT images generated by AI from UTE-MRI, without affecting the rest of the reconstructed image. Bronchiectasis of the left lower lobe (white arrowheads) was visually missed on **(A)** UTE-MRI leading to a false negative depiction of bronchiectasis, but clearer on **(B)** synthetic CT and confirmed on the **(C)** corresponding real CT scan.

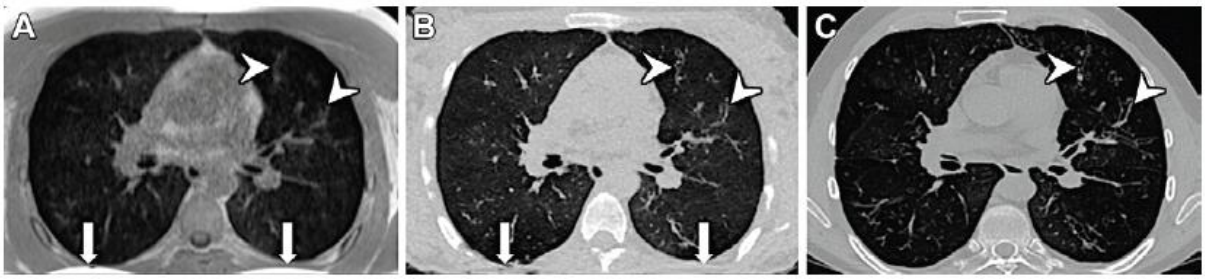

Figure 5. Example images of the lung in a 20-year-old female with Cystic Fibrosis from the external validation dataset. On **(A)** Lung MRI with ultrashort echo-times (UTE-MRI), aliasing of the mammary glands (white arrows) is visible partially outside the field-of-view due to the Cartesian k-space sampling in the anteroposterior axis of UTE-MRI. The aliasing artifacts were visible in the posterior extrapulmonary subcutaneous fat and did not impair the lung area on **(A)** UTE-MRI and were still noticeable on **(B)** synthetic CT images generated by AI from UTE-MRI. Note the presence of mild bronchiectasis in the left upper lobe (white arrowheads), with a conspicuous delineation on **(B)** synthetic CT and **(C)** corresponding real CT.

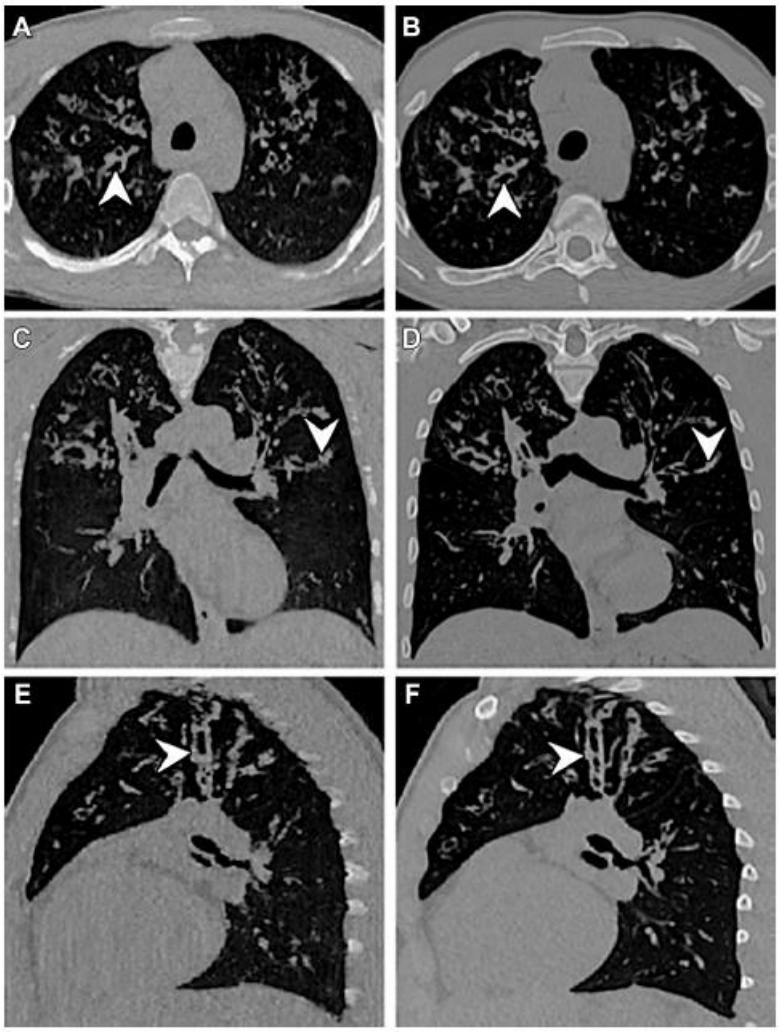

Figure 6. Example images of the lung in a 15-year-old male with Cystic Fibrosis from the external validation dataset. **(A, C, E)** show synthetic CT images generated by AI from lung MRI with ultrashort echo-times and **(B, D, F)** show corresponding real CT. **(A, B)** Axial, **(C, D)** coronal, and **(E, F)** sagittal reformations are possible owing to the 3D isotropic voxel at millimeter resolution. The Bhalla score was scored as 12 on both examinations.

SUPPLEMENTAL MATERIALS.

SUPPLEMENTAL METHODS.

Methods E1. Lung MRI with the spiral VIBE sequence.

3D isotropic millimetre imaging was used for human lung imaging because of the presence of fine structures such as bronchi and vessels. For this, a field-of-view (FOV) of (320x320mm) was used to cover the entire lung in the native coronal plane. A matrix resolution of 320x320 was used to reach a millimetre pixel resolution. The slice thickness was adapted to FOV size and matrix size to get an isotropic voxel dimension of (1x1x1mm). The number of slices had to be adapted to the patient size and ranged typically from 200 to 240. Patients were positioned in the supine position with arms along the body. To prevent aliasing due to the Cartesian sampling in the slice direction, imaging was performed in the coronal plane, covering the entire chest. The basic sequence parameters were as follows: TR/TE/flip angle=4.1ms/0.07ms/5°(4). The acquisition was performed during free breathing. To reduce respiratory-related artifacts, prospective gating was applied. For this purpose, navigator pulses were inserted into the sequence, from which the respiratory signal is extracted during scanning at functional residual capacity (end normal expiration). Once sufficient data was acquired in the end-expiratory state using a tolerance window of 40%, the scan stops automatically, and the collected data is used to reconstruct corresponding images. Acquisition time, therefore, depends on the respiratory pattern of the patient and ranged from 6 to 8 minutes. No injection of contrast agent or inhalation was used.

Methods E2. Generative Adversarial Networks framework.

To generate synthetic CT from lung MRI with ultrashort echo-times (UTE-MRI), a Training pipeline was implemented. To achieve high-resolution morphological imaging, CTs reconstructed with high frequency and pulmonary kernels were used for Training (Supplemental Table E2). Thus, all CT and MR 2D slices were first resampled in the axial axis on a common grid with a voxel size of $0.6\times0.6\times0.6mm^3$, and aligned each other using the EVolution algorithm (19). To speed up calculations and guide training, an automatic lung segmentation of the registered 2D CT images was computed using the U-net R-231 convolutional network (20). Since the lung MRI signal is not calibrated, CT intensities were

then cropped to [-1000; 2000] HU and rescaled to [- 1; 1]. Thus, MR intensity values were normalized using zero mean and unit variance. Herein, signal intensity values were cropped to $[-3\sigma; 3\sigma]$ to remove outliers, σ being the standard deviation of signal intensities, and rescale them to [-1; 1]. Then, GAN architecture was set up with Semantic Image Synthesis with Spatially-Adaptive Normalization (SPADE) (21). In SPADE, the learned affine parameters need to be spatially adaptive, which means that different scaling and bias are computed according to the semantic information learned during training. Using this method, previous experiments demonstrated the advantage of the SPADE method regarding both visual fidelity and alignment with input layouts (21). Moreover, the loss function was defined as a combination of feature-matching, perceptual, and contrastive (https://github.com/alexandonian/contrastive-feature-loss) (22). We use the Adam optimizer (23), a batch size of 1, a learning rate of 0.0001 and ReLU activation for the generator, and a learning rate of 0.0004 and LeakyReLU activation for the discriminator. The training process included 100 epochs and lasted around 400 hours on a 12GB NVIDIA RTX 2080 Ti. Finally, generating Synthetic CT from UTE with a trained model required around thirty seconds on one single GPU.

Methods E3. Quantitative assessment of imaging quality

Two readers with 4 years (J.R.) and 10 years (G.D.) of experience in chest imaging made the qualitative evaluation of imaging quality independently, and the mean of evaluations was chosen as the final result. All native UTE, CT and Synthetic CT data sets were anonymized, analyzed in random order, blinded from any clinical characteristics.

Regions of Interest (ROIs) of the same shape (circle) and size (56 mm^2) were drawn by G.D. and J.R. in the axial plane by using 3D Slicer 4.11, an open-source software. Region-of-interest (ROIs) were placed within the trachea, the right and left main bronchi and their signal values were averaged to calculate S_{air}. ROIs were placed within the pulmonary trunk, the right and left main pulmonary arteries and their signal values were averaged to calculate S_{vessel}.

Three axial sections were selected at the level of the crossing of the aorta, the carina and the inferior pulmonary veins. ROIs were placed at each location in the anterior part and posterior part of the right and left lung, at least 2 cm from the lung periphery and by careful avoidance of vessels. Data in the resulting 12 ROIs were averaged to calculate S_{lung}.

Apparent signal-to-noise ratio and apparent contrast-to-noise ratio were expressed as percentages and calculated as follows (1,2):

Signal-to-noise ratio = S_{lung} / S_{air}) x 100 (%)

Contrast-to-noise ratio = (S_{lung} – S_{airway}) / S_{vessel} x 100 (%).

Moreover, the standard deviation of air within the large tracheo-bronchial tree was chosen as a measurement of overall noise (24). Herein, three ROIs of were drawn within the mid portion of the trachea, the right mainstem and left mainstem bronchi, and averaged to get a single mean standard deviation value.

Supplemental Table E11 provides information on the inter observer reproducibility of independent measurements. Also, the junior reader assessed the evaluation twice to provide additional analysis of intra-observer reproducibility.

Methods E4. Qualitative assessment of imaging quality.

Two readers with 4 years (J.R.) and 10 years (G.D.) of experience in chest imaging made the qualitative evaluation of imaging quality independently Multiplanar reformations were allowed. All UTE, Synthetic CT and CT datasets were anonymized, analyzed in random order, blinded from any clinical characteristics. Then, J.R. and G.D. discussed the imaging data together during the same sessions. A third reader with 8 years of experience (I.B.) in thoracic imaging was consulted to reach a final qualitative value.

Per each lung, a qualitative score of imaging quality according to a scale adapted from Ohno et al (24) was performed on axial imaging slices at three levels: upper lung (above the cross of the aorta), mid lung (between the cross of the aorta and the superior pulmonary veins) and inferior lung (below the superior pulmonary veins). Moreover, the central lung was defined as the two third internal part of the lung, and the peripheral lung was defined as the third external part of the lung. For CT scans, a parenchymal windowing was chosen, with a width of 1500 HU and a level of -450 HU.

The visibility of central airways and central vessels was scored as follows: 0 = no depiction; 1 = depicted at segmental level; 2 = depicted at subsegmental level; 3 = depicted at sub-subsegmental level; 4 = depicted beyond sub-subsegmental level. The visibility of peripheral airways and peripheral vessels was scored as follows: 0 = not visible; 1 = visible. Then, the

total visibility score was calculated as the sum of central and peripheral scorings across the three upper, mid, and inferior lung levels. In the same regions, the sharpness of vessels and airways was score as follows: 0 = blurred; 1 = intermediate; 2 = sharp. Then, the total sharpness score was calculated as the sum of central and peripheral scorings across the three upper, mid, and inferior lung levels (Supplemental Table E3).

Finally, five artifacts were considered for analysis: beam hardening, motion, aliasing, shadowing, and streaks artifacts. The severity of these artifacts was scored between 0 and 3 as follows: 0 = no artifact; 1 = slight artefact without blurring of anatomical structures; 2 = moderate artefact with blurring of anatomical structures; 3 = severe artefact with no normal structure recognizable. No distinction was done between the central and peripheral lung herein. Thus, the total artefact score was calculated as the sum of scorings across three upper, mid, and inferior lung levels. (Supplemental Table E3).

Therefore, the (minimum to maximum) range of total visibility, sharpness, and artifact scores were (0 to 30), (0 to 24) and (0 to 45), respectively.

Supplemental Table E11 provides information on the inter observer reproducibility of independent evaluations. Also, the junior reader assessed the evaluation twice to provide additional analysis of intra-observer reproducibility.

Methods E5. Evaluation of CF structural alterations at the segmental level

To assess whether UTE-MRI and Synthetic CT may provide similar structural CF-related information to real CT, a senior reader with 10 years of experience in chest imaging (G.D.) and a published experience in both chest CT and MRI of cystic fibrosis completed the segmental analyses. For this task, the junior reader was left out of the evaluation, to keep him as naïve as possible from CF scoring with UTE. All UTE-MRI, Synthetic CT and CT datasets were anonymized and analyzed in random order, blinded from any clinical data. Multiplanar reformations were allowed. At the segmental level, the presence or absence of structural alterations was assessed using a binary scale (0= absent, 1=present).

Structural alterations were those necessary to complete the Bhalla scoring system (25) (see Supplemental Table E1) *i.e.*, peribronchial thickening, bronchiectasis, mucus plugs, bronchiolar impactions with the tree-in-bud pattern, collapse/consolidation, emphysema,

bulla, and sacculation/abscess. Definitions of structural abnormalities were those of the Fleishner Glossary of terms (26).

Methods E6. Evaluation of the Bhalla score.

Structural alterations were evaluated according to the original version of the standard Bhalla scoring system (25) (see Supplemental Table E4). Lung structural abnormalities, i.e., peribronchial thickening, bronchiectasis, mucus plugs, bronchiolar impactions with the tree-in-bud pattern, collapse/consolidation, emphysema, bulla, and sacculation/abscess. Definitions of structural abnormalities were those of the Fleishner Glossary of terms (26).

According to the scoring system, bronchiectasis and peribronchial thickening were assigned a severity score as follows:

- Bronchiectasis severity: Absent (0), luminal diameter slightly greater than the diameter of the adjacent vessel (Mild (1)), luminal diameter 2–3× the adjacent vessel (Moderate (2)), luminal diameter > 3× the adjacent vessel (Severe (3)).

- Peribronchial thickening severity: Absent (0), wall thickness equal to the diameter of adjacent vessel (Mild (1)), wall thickness 2× the adjacent vessel (Moderate (2)), wall thickness > 2× the adjacent vessel (Severe (3)).

The bronchial generation involved by bronchiectasis or mucus impactions was scored as follows: Absent (0); up to the 4^{th} generation (1); up to the fifth generation (2); up to the 6^{th} generation and above (3).

Then, the score is calculated as follows: Total Bhalla score = 25 – (sum of scorings).

The evaluations were performed independently by GD and JR, twice, to assess the intra and interobserver reproducibility (Supplemental Table E10).

SUPPLEMENTAL TABLES.

Table E1. CT characteristics in Training, Test and External datasets.

Groups	Machine Brand	Kernel	kV	mAs	Slice thickness (mm)	CTDIvol (mGy)	DLP (mGy.cm)
Training	GE Revolution® (Site1, n=62)	LUNG (n=51); I50f3 (n=9) ; BI57d3 (n=2)	(80-120)	(9-2195)	(0.625-1.25)	(0.8-6)	(15-250)
	Somatom Definition 64® (Site1, n=4)	B70f (n=2); B50f (n=1) ; I50f3 (n=1)					
	Somatom Force® (Site1, n=16)	BI57d3 (n=10); BI57d5 (n=3); LUNG (n=3)					
Test	GE Revolution® (Site1, n=10)	LUNG (n=9); BI57d3 (n=1)	(80-140)	(9-808)	(0.625-1)	(0.6-7)	(13-245)
	Somatom Definition 64® (Site1, n=1)	B50f (n=1)					
	Somatom Force® (Site1, n=17)	BI57d3 (n=2); BI57d5 (n=15)					
External validation	GE Revolution® (Site2, n=6; Site3, n=8 ;)	LUNG (n=14)	(70-150)	(9-1610)	(0.625-1.25)	(0.8-7.5)	(16-260)
	GE OptimaCT660® (Site 4, n=11; Site2, n=3)	LUNG (n=14)					
	Somatom Definition 64® (Site5, n=11; Site6, n=3)	I50f3 (n=4); BI57d3 (n=7); B70f (n=3)					
	Somatom Force® (Site5, n=4)	BI57d3 (n=4)					

Legend: Site1=University Hospital of Bordeaux; Site2=University Hospital of Grenoble; Site3=University Hospital of Marseille; Site4=University Hospital of Nantes; Site5=University Hospital of Lille; Site6=University Hospital of Tours; kV=kilovoltage, mAs=milliampere second; CTDIvol=volumic CT dose index; DLP = dose length product; data between parentheses are the (minimum-maximum) range of values.

Table E2. Patients' characteristics of the Training dataset

	Training dataset (n=82)
Age (years)	21 ± 11
Sex	
Male	42 (51%)
Female	40 (49%)
BMI ($kg.m^{-2}$)	18 ± 3
DeltaF508 mutation	
Homozygous	44 (53%)
Heterozygous	32 (40%)
Other mutation	6 (7%)
Pulmonary function test	
FEV1 (%)	79 ± 25
FEV1/FVC	72 ± 11
RV (%)	134 ± 42
TLC (%)	103 ± 13
CT	
Bhalla score	14.9 ± 4.6

Note. —Data are mean ± standard deviation for continuous data and number of patients with percentages in parentheses for categorical data. BMI = body mass index, FEV1 = forced expiratory volume in 1 second, FVC = forced vital capacity, RV = residual volume, TLC = total lung capacity

Table E3. Template of visibility, sharpness and artefact scores.

Visibility score template	Range of scorings
Central lung	
Upper lung	
Vessel	(0 - 4)*
Airways	(0 - 4)*
Mid Lung	
Vessel	(0 - 4)*
Airways	(0 - 4)*
Bottom lung	
Vessel	(0 - 4)*
Airways	(0 - 4)*
Peripheral lung	
Upper lung	
Vessel	(0 - 1)**
Airways	(0 - 1)**
Mid lung	
Vessel	(0 - 1)**
Airways	(0 - 1)**
Bottom lung	
Vessel	(0 - 1)**
Airways	(0 - 1)**
Total visibility score	
Sum of visibility scorings	(0 - 30)

Sharpness score template	Range of scorings
Central lung	
Upper lung	
Vessel	(0 - 2)†
Airways	(0 - 2)†
Mid Lung	
Vessel	(0 - 2)†
Airways	(0 - 2)†
Bottom lung	
Vessel	(0 - 2)†
Airways	(0 - 2)†
Peripheral lung	
Upper lung	
Vessel	(0 - 2)†
Airways	(0 - 2)†
Mid lung	
Vessel	(0 - 2)†
Airways	(0 - 2)†
Bottom lung	
Vessel	(0 - 2)†
Airways	(0 - 2)†
Total visibility score	
Sum of sharpness scorings	(0 - 24)

Artefact score template	Range of scorings
Beam hardening	
Upper lung	(0 - 3)‡
Mid lung	(0 - 3)‡
Bottom lung	(0 - 3)‡
Motion	
Upper lung	(0 - 3)‡
Mid lung	(0 - 3)‡
Bottom lung	(0 - 3)‡
Aliasing	
Upper lung	(0 - 3)‡
Mid lung	(0 - 3)‡
Bottom lung	(0 - 3)‡
Shadowing	
Upper lung	(0 - 3)‡
Mid lung	(0 - 3)‡
Bottom lung	(0 - 3)‡
Streaks	
Upper lung	(0 - 3)‡
Mid lung	(0 - 3)‡
Bottom lung	(0 - 3)‡
Total artefact score	
Sum of artefact scorings	(0 - 45)

Note: * for visibility scoring in the central lung, the scores were as follows: 0 = no depiction; 1 = depicted at segmental level; 2 = depicted at subsegmental level; 3 = depicted at sub-subsegmental level; 4 = depicted beyond sub-subsegmental level. Then, the total sharpness score was calculated as the sum of central and peripheral scorings across the three upper, mid, and inferior lung levels (Supplemental Table E).

** for visibility scoring in the peripheral lung, the scores were as follows: 0 = not visible; 1 = visible.

† for sharpness scorings in the central and peripheral lung, the scores were as follows: 0 = blurred; 1 = intermediate; 2 = sharp.

‡ for artefact scorings, the scores were as follows: 0 = no artifact; 1 = slight artefact without blurring of anatomical structures; 2 = moderate artefact with blurring of anatomical structures; 3 = severe artefact with no normal structure recognizable.

Table E4. Bhalla score template.

Category	Scores			
	0	1	2	3
Severity of bronchiectasis	Absent	Mild	Moderate	Severe
Peribronchial thickening	Absent	Mild	Moderate	Severe
Extent of bronchiectasis (no. of segments)	Absent	1–5	6–9	>9
Extent of mucus plugging (no. of segments)	Absent	1–5	6–9	>9
Sacculatoins or abscesses (no. of segments)	Absent	1–5	6–9	>9
Generation of bronchial divisions involved (bronchiectasis/plugging)	Absent	Up to 4th generation	Up to 5th generation	Up to 6th generation and distal
No. of bullae	Absent	Unilateral (not > 4)	Bilateral (not > 4)	>4
Emphysema (no. of segments)	Absent	1–5	>5	
Collapse/consolidation	Absent	Subsegmental	Segmental/lobar	

Note: The template is adapted from Bhalla et al, Radiology 1991. The score is calculated as follows: Total Bhalla score = 25 – (sum of scorings).

When present, bronchiectasis and peribronchial thickening were assigned a severity score as follows:

Bronchiectasis severity: Absent (0), luminal diameter slightly greater than diameter of adjacent vessel (Mild (1)), luminal diameter 2–3× the adjacent vessel (Moderate (2)), liminal diameter > 3× the adjacent vessel (Severe (3)).

Peribronchial thickening severity: Absent (0), wall thickness equal to the diameter of adjacent vessel (Mild (1)), wall thickness 2× the adjacent vessel (Moderate (2)), wall thickness > 2× the adjacent vessel (Severe (3)).

The authors acknowledge that other versions of the Bhalla score have been proposed in the literature, out of the scope of this study.

Table E5. Comparison of quality imaging in the Test and the External datasets.

Test dataset (n = 28)	**UTE-MRI**	**Synthetic CT**	**CT**
Quantitative analysis			
Contrast-to-noise ratio	p1 < .001	p2 = 0.46	p3 < .001
Signal-to-noise-ratio	p1 = 0.88	p2 = 0.96	p3 = 0.76
Overall Noise	p1 = 0.75	p2 < .001	p3 < .001
Qualitative analysis			
Total visibility score	p1 < .001	p2 = 0.89	p3 < .001
Total sharpness score	p1 < .001	p2 = 0.92	p3 < .001
Total artifact score	p1 < .001	p2 < .001	p3 = 0.26

External dataset (n=46 patients)	**UTE-MRI**	**Synthetic CT**	**CT**
Quantitative analysis			
Contrast-to-noise ratio	p1 < .001	p2 = 0.34	p3 < .001
Signal-to-noise ratio	p1 = 0.76	p2 = 0.87	p3 = 0.85
Overall Noise	p1 = 0.09	p2 < .001	p3 < .001
Qualitative analysis			
Total visibility score	p1 < .001	p2 = 0.43	p3 < .001
Total sharpness score	p1 < .001	p2 = 0.29	p3 < .001
Total artifacts score	p1 < .001	p2 < 0.001	p3 < 0.08

Note: data are p-value calculations between paired medians according to a post-hoc Dunn's test. P1 indicate the p-value of comparison between UTE-MRI and Synthetic CT, p2 between Synthetic CT and CT, and p3 between CT and UTE-MRI.

Table E6. Qualitative analysis of imaging quality in the Test dataset, with details on visual scorings.

	UTE-MRI		Synthetic CT		CT		p-value
Qualitative analysis							
Central lung							
Vessel visibility	12	[12, 12]	12	[12, 12]	12	[12, 12]	0.37
Airway visibility	**10.5****,†	[9, 12]	12	[12, 12]	12	[12, 12]	<0.001
Vessel sharpness	**3****,†	[3, 3]	6	[6, 6]	6	[6, 6]	<0.001
Airway sharpness	**3****,†	[3, 3]	6	[6, 6]	6	[6, 6]	<0.001
Peripheral lung							
Vessel visibility	3	[3, 3]	3	[3, 3]	3	[3, 3]	<0.001
Airway visibility	2	[0, 3]	2	[0, 3]	2	[0, 3]	0.15
Vessel sharpness	**3****,†	[3, 3]	6	[6, 6]	6	[6, 6]	<0.001
Airway sharpness	**2****,†	[0, 3]	4	[0, 6]	4	[0, 6]	<0.001
Artifacts							
Beam hardening	0	[0, 0]	0	[0, 0]	**1***,†	[0, 1]	<0.001
Motion	**0.5****,†	[0, 1]	0	[0, 0]	0	[0, 0]	<0.001
Aliasing	0	[0, 0]	0	[0, 0]	0	[0, 0]	<0.001
Shadowing	0	[0, 0]	0	[0, 0]	0	[0, 0]	<0.001
Streak	0	[0, 0]	0	[0, 0]	0	[0, 0]	<0.001

Note: per each patient, in the central lung, the visibility and sharpness scorings have a (minimum to maximum) range of values of (0 to 12) and (0 to 6), respectively; in the peripheral lung, the visibility and sharpness scorings have a (minimum to maximum) range of values of (0 to 3) and (0 to 6), respectively. The artifacts scorings have a (minimum to maximum) range of values of (0 to 9).

Data are median with [interquartile range]. Comparison of medians were assessed with Friedman test and post hoc Dunn's test. * indicates a p-value ≤ .001 between CT and Synthetic CT, ** a p-value ≤ .001 between UTE-MRI and Synthetic CT and † p-value ≤ .001 between UTE-MRI and CT.

Legends: UTE-MRI=lung MRI with ultrashort echo times

Table E7. Qualitative analysis of imaging quality in the External dataset, with details on visual scorings.

	UTE-MRI		Synthetic CT		CT		p-value
Qualitative analysis							
Central lung							
Vessel visibility	12	[12, 12]	12	[12, 12]	12	[12, 12]	<0.001
Airway visibility	**12****,†	[10, 12]	12	[12, 12]	12	[12, 12]	<0.001
Vessel sharpness	**3****,†	[3, 3]	6	[6, 6]	6	[6, 6]	<0.001
Airway sharpness	**3****,†	[3, 3]	6	[6, 6]	6	[6, 6]	<0.001
Peripheral lung							
Vessel visibility	3	[3, 3]	3	[3, 3]	3	[3, 3]	<0.001
Airway visibility	2	[1, 2]	2	[1, 3]	2	[0, 3]	0.003
Vessel sharpness	**3****,†	[3, 3]	6	[6, 6]	6	[6, 6]	<0.001
Airway sharpness	**2****,†	[1, 3]	4	[2, 6]	**4**	[0, 6]	<0.001
Artifacts							
Beam hardening	0	[0, 0]	0	[0, 0]	**0.5***,†	[0, 1]	<0.001
Motion	**1****,†	[0, 3]	0	[0, 0]	0	[0, 0]	<0.001
Aliasing	0	[0, 0]	0	[0, 0]	0	[0, 0]	0.37
Shadowing	0	[0, 0]	0	[0, 0]	0	[0, 0]	<0.001
Streak	0	[0, 0]	0	[0, 0]	0	[0, 0]	<0.001

Note: per each patient, in the central lung, the visibility and sharpness scorings have a (minimum to maximum) range of values of (0 to 12) and (0 to 6), respectively; in the peripheral lung, the visibility and sharpness scorings have a (minimum to maximum) range of values of (0 to 3) and (0 to 6), respectively. The artifacts scorings have a (minimum to maximum) range of values of (0 to 9).

Data are median with [interquartile range]. Comparison of medians were assessed with Friedman test and post hoc Dunn's test. * indicates a p-value ≤ .001 between CT and Synthetic CT, ** a p-value ≤ .001 between UTE-MRI and Synthetic CT and † p-value ≤ .001 between UTE-MRI and CT.

Legends: UTE-MRI=lung MRI with ultrashort echo times

Table E8. Comparison of imaging quality between the Test dataset and External dataset.

	Test dataset (n = 28)		External dataset (n = 46)		p-value
Quantitative analysis					
UTE-MRI					
Contrast-to-noise ratio	9.3	[6.6, 35]	13	[7, 20]	0.81
Signal-to-noise ratio	88	[87, 91]	90	[88, 93]	0.10
Overall noise	22	[18, 31]	21.5	[14, 45]	0.67
Synthetic CT					
Contrast-to-noise ratio	303	[221, 382]	364	[258, 448]	0.22
Signal-to-noise ratio	88	[84, 92]	90	[88, 92]	0.35
Overall noise	26	[22, 30]	23	[18, 28]	0.12
CT					
Contrast-to-noise ratio	247	[210, 323]	245	[155, 324]	0.29
Signal-to-noise ratio	88	[86, 91]	90	[87, 92]	0.14
Overall noise	42	[32, 50]	39	[20, 46]	0.08
Qualitative analysis					
UTE-MRI					
Total visibility score	27	[24, 29]	28	[26, 29]	0.10
Total sharpness score	10.5	[9, 12]	11	[9, 18]	0.32
Total artefact score	0.5	[0, 1]	1	[0, 3]	**0.03**
Synthetic CT					
Total visibility score	29	[27, 30]	29	[28, 30]	0.74
Total sharpness score	22	[18, 24]	22	[20, 24]	0.62
Total artefact score	0	[0, 0]	0	[0, 0]	0.43
CT					
Total visibility score	29	[27, 30]	29	[28, 30]	0.90
Total sharpness score	22	[18, 24]	22	[18, 24]	0.35
Total artefact score	1	[0, 1]	0.5	[0, 1]	0.83

Note: data are median with [interquartile range] for qualitative and quantitative measurements. Comparison of independent medians are done with the Mann-Whitney test.

Legends: UTE-MRI=lung MRI with ultrashort echo times

Table E9. Contingency tables to detect CF-related structural alterations at the segmental level.

CT vs Synthetic CT in the Test dataset (n=504 segments in 28 patients)

	WT	BD	MPc	MPp	Cons	Emph	Sacc	Bulla
TP	266	236	130	177	12	0	3	0
TN	211	247	372	323	492	504	501	504
FP	16	5	1	2	0	0	0	0
FN	11	16	1	2	0	0	0	0

CT vs UTE-MRI in the Test dataset (n=504 segments in 28 patients)

	WT	BD	MPc	MPp	Cons	Emph	Sacc	Bulla
TP	231	197	126	170	10	0	3	0
TN	200	249	361	311	492	504	501	504
FP	27	3	12	14	0	0	0	0
FN	46	55	5	9	2	0	0	0

CT vs Synthetic CT in the External dataset (n=828 segments in 46 patients)

	WT	BD	MPc	MPp	Cons	Emph	Sacc	Bulla
TP	693	671	487	505	28	0	9	1
TN	123	151	318	304	799	828	819	826
FP	2	0	19	9	0	0	0	0
FN	10	6	4	10	1	0	0	1

CT vs UTE-MRI in the External dataset (n=828 segments in 46 patients)

	WT	BD	MPc	MPp	Cons	Emph	Sacc	Bulla
TP	667	624	453	468	28	0	9	1
TN	119	148	320	296	797	828	817	826
FP	6	3	17	17	2	0	1	0
FN	36	53	38	47	1	0	1	1

Note: data are absolute number of segments with or without a given structural alteration, according to the senior reader, with CT as reference.

Legends: UTE-MRI=lung MRI with ultrashort echo-times; CT=real computed tomography; TP=true positive depiction; TN=true negative depiction; FP=false positive depiction; FN=false negative depiction; WT=peribronchial wall thickening; BD=bronchial dilatation; MPc=central mucus plugging; MPp=peripheral mucus pluggning with a “tree-in-bud” appearance; Cons=collapse/consolidation; Emph=emphysema; Sacc=Sacculation/Abscess.

Table E10. Correlations between Bhalla scores and pulmonary function tests in the Test dataset.

Bhalla score (n=28)	Pulmonary function tests					
	FEV1%p		FEV1/FVC		RV%p	
	r	p-value	r	p-value	r	p-value
Senior reader						
CT	0.62	**<0.001**	0.52	**0.005**	-0.42	**0.02**
Synthetic CT	0.62	**<0.001**	0.53	**0.004**	-0.43	**0.02**
UTE-MRI	0.60	**<0.001**	0.45	**0.01**	-0.39	**0.03**
Junior reader						
CT	0.62	**<0.001**	0.49	**0.008**	-0.34	0.07
Synthetic CT	0.65	**<0.001**	0.48	**0.009**	-0.34	0.07
UTE-MRI	0.46	**0.01**	0.33	0.08	-0.20	0.30

Note: data are r correlation coefficients of Pearson obtained from the first reading sessions per each reader. A p-value inferior to 0.05 was considered significant.

Legends: UTE-MRI=lung MRI with ultrashort echo-times; CT=real computed tomography; FEV1=forced expiratory volume in 1 second; FVC=forced vital capacity; RV=residual volume; %p=percentage predicted.

Table E11. Intra and inter observer reproducibility of Bhalla scorings.

Test dataset (n = 28)		**Junior reader**		**Senior reader**	
		UTE-MRI1	UTE-MRI2	UTE-MRI1	UTE-MRI2
Junior reader	UTE-MRI1	NA	-	-	-
	UTE-MRI2	**0.90**	NA	-	-
Senior reader	UTE-MRI1	0.72	0.77	NA	-
	UTE-MRI2	0.77	0.77	**0.98**	NA

		Junior reader		**Senior reader**	
		Synthetic CT1	Synthetic CT2	Synthetic CT1	Synthetic CT2
Junior reader	Synthetic CT1	NA	-	-	-
	Synthetic CT2	**0.95**	NA	-	-
Senior reader	Synthetic CT1	0.91	0.92	NA	-
	Synthetic CT2	0.90	0.90	**0.99**	NA

		Junior reader		**Senior reader**	
		CT1	CT2	CT1	CT2
Junior reader	CT1	NA	-	-	-
	CT2	**0.98**	NA	-	-
Senior reader	CT1	0.93	0.94	NA	-
	CT2	0.93	0.94	**0.99**	NA

External dataset (n = 46)		**Junior reader**		**Senior reader**	
		UTE-MRI1	UTE-MRI2	UTE-MRI1	UTE-MRI2
Junior reader	UTE-MRI1	NA	-	-	-
	UTE-MRI2	**0.90**	NA	-	-
Senior reader	UTE-MRI1	0.75	0.74	NA	-
	UTE-MRI2	0.66	0.65	**0.99**	NA

		Junior reader		**Senior reader**	
		Synthetic CT1	Synthetic CT2	Synthetic CT1	Synthetic CT2
Junior reader	Synthetic CT1	NA	-	-	-
	Synthetic CT2	**0.98**	NA	-	-
Senior reader	Synthetic CT1	0.92	0.90	NA	-
	Synthetic CT2	0.91	0.90	**0.98**	NA

		Junior reader		**Senior reader**	
		CT1	CT2	CT1	CT2
Junior reader	CT1	NA	-	-	-
	CT2	**0.98**	NA	-	-
Senior reader	CT1	0.93	0.94	NA	-
	CT2	0.93	0.94	**0.99**	NA

Note: Data are intraclass correlation coefficients. The range of the Bhalla score was (0 – 25). UTE-MRIx = lung MRI with ultrashort echotimes at the xth reading session; CTx = CT at the xth reading session; NA=not applicable.

Table E12. Inter and intra-observer reproducibility of imaging quality assessments.

	Test dataset (n = 28)				External dataset (n = 46)			
	ICC inter	95% CI	**ICC intra**	95% CI	**ICC inter**	95% CI	**ICC intra**	95% CI
Contrast to-noise ratio								
UTE-MRI	0.99	[0.98, 0.99]	0.99	[0.98, 0.99]	0.97	[0.95, 0.98]	0.99	[0.98, 0.99]
Synthetic CT	0.96	[0.91, 0.98]	0.99	[0.97, 0.99]	0.95	[0.91, 0.97]	0.99	[0.98, 0.99]
CT	0.98	[0.97, 0.99]	0.99	[0.98, 0.99]	0.96	[0.92, 0.97]	0.99	[0.98, 0.99]
Signal-to-noise ratio								
UTE-MRI	0.94	[0.84, 0.97]	0.93	[0.85, 0.96]	0.94	[0.89, 0.96]	0.93	[0.87, 0.96]
Synthetic CT	0.95	[0.82, 0.98]	0.95	[0.89, 0.97]	0.92	[0.87, 0.95]	0.92	[0.84, 0.95]
CT	0.97	[0.94, 0.98]	0.96	[0.91, 0.98]	0.96	[0.93, 0.97]	0.93	[0.85, 0.96]
Overall noise								
UTE-MRI	0.92	[0.85, 0.93]	0.99	[0.98, 0.99]	0.91	[0.85, 0.95]	0.96	[0.93, 0.97]
Synthetic CT	0.97	[0.94, 0.98]	0.97	[0.95, 0.99]	0.92	[0.86, 0.95]	0.96	[0.93, 0.97]
CT	0.99	[0.98, 0.99]	0.99	[0.98, 0.99]	0.98	[0.97, 0.99]	0.99	[0.98, 0.99]
Total visibility score								
UTE-MRI	0.97	[0.94, 0.98]	0.93	[0.85, 0.97]	0.98	[0.96, 0.98]	0.98	[0.97, 0.99]
Synthetic CT	0.94	[0.89, 0.97]	0.94	[0.87, 0.97]	0.95	[0.91, 0.97]	0.90	[0.84, 0.94]
CT	0.94	[0.89, 0.97]	0.95	[0.90, 0.97]	0.98	[0.97, 0.99]	0.98	[0.97, 0.99]
Total sharpness score								
UTE-MRI	0.98	[0.97, 0.99]	0.99	[0.99, 0.99]	0.98	[0.96, 0.98]	0.98	[0.97, 0.99]
Synthetic CT	0.99	[0.98, 0.99]	0.99	[0.99, 0.99]	0.98	[0.97, 0.99]	0.99	[0.99, 0.99]
CT	0.99	[0.99, 0.99]	0.99	[0.99, 0.99]	0.99	[0.99, 0.99]	0.99	[0.99, 0.99]
Total artefact score								
UTE-MRI	0.97	[0.95, 0.99]	0.98	[0.97, 0.99]	0.97	[0.95, 0.98]	0.97	[0.96, 0.98]
Synthetic CT	NA	NA	NA	NA	1	[1, 1]	1	[1, 1]
CT	0.97	[0.95, 0.99]	0.98	[0.97, 0.99]	0.98	[0.97, 0.99]	0.98	[0.97, 0.99]

Note: data are intraclass correlation coefficients with 95% confidence interval (CI). The range of the total visibility score, sharpness score and artefact score were (0 – 30), (0 – 24) and (0 – 35), respectively.

ICC inter = intraclass correlation coefficient of inter-observer reproducibility; ICC intra = intraclass correlation coefficient of intra-observer reproducibility; UTE-MRI = lung MRI with ultrashort echo times.