



OPEN ACCESS

EDITED BY

Yogesh Kumar,
Pandit Deendayal Energy University, India

REVIEWED BY

Orso Pusterla,
University Hospital of Basel, Switzerland
Apeksha Koul,
Punjabi University, India
Nandini Modi,
Pandit Deendayal Energy University, India

*CORRESPONDENCE

Urs Eisenmann
✉ urs.eisenmann@med.uni-heidelberg.de

RECEIVED 15 March 2024

ACCEPTED 21 October 2024

PUBLISHED 12 November 2024

CITATION

Ringwald FG, Wucherpfennig L, Hagen N, Mücke J, Kaletta S, Eichinger M, Stahl M, Triphan SMF, Leutz-Schmidt P, Gestewitz S, Graeber SY, Kauczor H-U, Alrajab A, Schenk J-P, Sommerburg O, Mall MA, Knaup P, Wielpütz MO and Eisenmann U (2024) Automated lung segmentation on chest MRI in children with cystic fibrosis. *Front. Med.* 11:1401473. doi: 10.3389/fmed.2024.1401473

COPYRIGHT

© 2024 Ringwald, Wucherpfennig, Hagen, Mücke, Kaletta, Eichinger, Stahl, Triphan, Leutz-Schmidt, Gestewitz, Graeber, Kauczor, Alrajab, Schenk, Sommerburg, Mall, Knaup, Wielpütz and Eisenmann. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Automated lung segmentation on chest MRI in children with cystic fibrosis

Friedemann G. Ringwald^{1,2}, Lena Wucherpfennig^{2,3,4}, Niclas Hagen^{1,2}, Jonas Mücke¹, Sebastian Kaletta¹, Monika Eichinger^{2,3,4}, Mirjam Stahl^{2,5,6,7}, Simon M. F. Triphan^{2,3,4}, Patricia Leutz-Schmidt^{2,3,4}, Sonja Gestewitz^{2,3,4}, Simon Y. Graeber^{2,5,6,7}, Hans-Ulrich Kauczor^{2,3,4}, Abdulsattar Alrajab³, Jens-Peter Schenk³, Olaf Sommerburg^{2,8,9}, Marcus A. Mall^{2,5,6,7}, Petra Knaup^{1,2}, Mark O. Wielpütz^{2,3,4} and Urs Eisenmann^{1,2*}

¹Institute of Medical Informatics, Heidelberg University, Heidelberg, Germany, ²Translational Lung Research Center Heidelberg (TLRC), German Center for Lung Research (DZL), Heidelberg, Germany, ³Department of Diagnostic and Interventional Radiology, University Hospital Heidelberg, Heidelberg, Germany, ⁴Department of Diagnostic and Interventional Radiology with Nuclear Medicine, Thoraxklinik at University Hospital Heidelberg, Heidelberg, Germany, ⁵Department of Pediatric Respiratory Medicine, Immunology and Critical Care Medicine, Charité-Universitätsmedizin Berlin, Berlin, Germany, ⁶German Center for Lung Research (DZL), Associated Partner Site, Berlin, Germany, ⁷Berlin Institute of Health (BIH) at Charité-Universitätsmedizin Berlin, Berlin, Germany, ⁸Division of Pediatric Pulmonology & Allergy and Cystic Fibrosis Center, Department of Pediatrics, University Hospital Heidelberg, Heidelberg, Germany, ⁹Department of Translational Pulmonology, University Hospital Heidelberg, Heidelberg, Germany

Introduction: Segmentation of lung structures in medical imaging is crucial for the application of automated post-processing steps on lung diseases like cystic fibrosis (CF). Recently, machine learning methods, particularly neural networks, have demonstrated remarkable improvements, often outperforming conventional segmentation methods. Nonetheless, challenges still remain when attempting to segment various imaging modalities and diseases, especially when the visual characteristics of pathologic findings significantly deviate from healthy tissue.

Methods: Our study focuses on imaging of pediatric CF patients [mean age, standard deviation (7.50 ± 4.6)], utilizing deep learning-based methods for automated lung segmentation from chest magnetic resonance imaging (MRI). A total of 165 standardized annual surveillance MRI scans from 84 patients with CF were segmented using the nnU-Net framework. Patient cases represented a range of disease severities and ages. The nnU-Net was trained and evaluated on three MRI sequences (BLADE, VIBE, and HASTE), which are highly relevant for the evaluation of CF induced lung changes. We utilized 40 cases for training per sequence, and tested with 15 cases per sequence, using the Sørensen-Dice-Score, Pearson's correlation coefficient (r), a segmentation questionnaire, and slice-based analysis.

Results: The results demonstrated a high level of segmentation performance across all sequences, with only minor differences observed in the mean Dice coefficient: BLADE (0.96 ± 0.05), VIBE (0.96 ± 0.04), and HASTE (0.95 ± 0.05). Additionally, the segmentation quality was consistent across different disease severities, patient ages, and sizes. Manual evaluation identified specific challenges, such as incomplete segmentations near the diaphragm and dorsal

regions. Validation on a separate, external dataset of nine toddlers (2–24 months) demonstrated generalizability of the trained model achieving a Dice coefficient of 0.85 ± 0.03 .

Discussion and conclusion: Overall, our study demonstrates the feasibility and effectiveness of using nnU-Net for automated segmentation of lung halves in pediatric CF patients, showing promising directions for advanced image analysis techniques to assist in clinical decision-making and monitoring of CF lung disease progression. Despite these achievements, further improvements are needed to address specific segmentation challenges and enhance generalizability.

KEYWORDS

deep learning, magnetic resonance imaging, cystic fibrosis, lung segmentation, pediatric

1 Introduction

Cystic fibrosis (CF) is an inherited multi-organ disease, which largely affects the lungs. Repeated bacterial infections and inflammation can result in lung damage, causing most of the morbidity and mortality seen in CF (1, 2). The early detection and monitoring of CF-related lung disease is a prerequisite for optimized care and improved long-term outcomes (3–7).

Recently, chest magnetic resonance imaging (MRI), a radiation-free modality, has shown great promise in assessing structural and functional CF lung abnormalities. Studies have shown chest MRI can detect changes as early as in infancy, and is capable of monitoring disease progression and therapeutic response throughout adulthood (7–14).

To semi-quantitatively assess the severity of lung abnormalities in CF patients, a morpho-functional chest MRI scoring system, also referred to as the Eichinger Score, was developed in 2012 (15). This scoring system includes items for morphological lung abnormalities, as well as perfusion abnormalities (8–11, 15, 16). To automate this scoring process, a critical step is automating the lung segmentation process.

In medical imaging, segmentation refers to identifying an organ or specific tissue of interest by extracting the boundaries and the inner region. This process allows for downstream analysis and extracting important quantitative information within that region. Precise segmentation may support accurate decisions on diagnosis, treatment plans, disease monitoring, and guiding of interventions (17). In the last decade, automated segmentation methods improved in performance and precision, resulting in the possibility of fully automated segmentation in different medical disciplines and imaging modalities (18). Machine learning methods, particularly neural networks, have demonstrated remarkable performance, often outperforming conventional methods, especially when analyzing large datasets (19–21). The nnU-Net, an advanced deep learning framework tailored for medical applications, stands out in its performance (22). It permits the training of networks to perform semantic segmentation with high accuracy and performance, eliminating the need for numerous configuration steps due to its self-configuring training parameters and layer settings. However, difficulties arise when attempting to adapt the nnU-Net to a variety of imaging modalities and diseases. This is particularly challenging when the visual characteristics of pathologic

findings deviate significantly from healthy tissue, indicating a change in tissue composition within the same organ (23).

Automated lung segmentation in MR images, especially in the CF population, also have inherent challenges. In MRI, difficulties arise due to the limited spatial resolution and the low contrast between the lungs and the adjacent tissue. In CF patients, breathing artifacts, most notably in young children; cardiac pulsation artifacts; chest growth in children, lung abnormalities displacing air contents, and the deformation associated with disease progression, all contribute to the complexity of the segmentation task (8, 24).

Despite these challenges, many studies are beginning to show promising results incorporating neural networks to automate MRI lung segmentation, even in different underlying pathologies, replacing conventional segmentation approaches (25, 26). Zha et al. applied convolutional neural networks (CNNs) on 3D radial ultra-short echo-times (UTE) oxygen-enhanced MRI in a dataset of 45 subjects (age 10+ with CF, asthma, or healthy) and achieved Dice coefficients of 0.97 and 0.96 for the right and left lung, respectively (27). Furthermore, researchers tested other MRI sequences, such as fast UTE with stack-of-spirals trajectory and matrix pencil decomposition MRI, in CF patients (age 5+) yielding Dice coefficients of 0.96 for children and 0.89 for adults (28, 29).

Notably, Astley et al. tested 2D and 3D nnU-Nets for lung segmentation of patients with varying pulmonary pathologies. In their patient cohort (median age 34 yrs.), analysis of a dataset comprising 809 spoiled-gradient-recalled and UTE MRI scans, even across different vendors, demonstrated a remarkable performance, reaching a median Dice coefficient of 0.96 internally and 0.97 on an external test set (30). To enhance the accuracy of automated lung segmentation, by inclusion of artificially generated images with consolidations, Cristoso et al. reached a Dice coefficient of 0.94 on a cohort of healthy volunteers and patients (31). In a 2023 study of neonates, either healthy or suffering from bronchopulmonary dysplasia, the authors employed CNNs for lung segmentation on quiet-breathing MRI and achieved a Dice coefficient of 0.908 on an internal test set and 0.88 on an independent test set (32). Most recently, a new approach for lung segmentation on healthy adults using thresholding and clustering on an enhanced deep-inspiration-breath-hold reached a Dice coefficient of 0.94 (33, 34). A high benchmark for lung lobe segmentation using pseudo-MRI images derived from CT and three concatenated CNNs achieved a Dice coefficient of 0.95 on a dataset of 100 CF patients over the age of 4.7 years old (35).

To the best of our knowledge, we are the first to demonstrate pediatric lung half segmentations for patients across the entire pediatric age range with different stages of cystic fibrosis using chest MRI on the commonly used sequences BLADE, VIBE, and HASTE. We selected a total of 165 MRI examinations from 84 patients in our internal monocentric CF database. This database contains 1,312 highly standardized annual surveillance MRIs, acquired over more than a decade from 266 patients. Segmentations were created manually by three observers.

2 Materials and methods

2.1 Study population

This ongoing prospective longitudinal observational study (clinicaltrials.gov identifiers NCT00760071, NCT02270476) was approved by the institutional ethics committee and informed written consent was obtained from the parents or legal guardians of all patients. The CF diagnosis was confirmed by increased sweat chloride (Cl⁻) concentrations (≥ 60 mmol/L) and cystic fibrosis transmembrane conductance regulator (CFTR) mutation analysis. In pancreatic-sufficient patients with borderline sweat test results (sweat Cl⁻ 30–60 mmol/L), the diagnosis was further supported by assessing CFTR function in rectal biopsies, as previously described (36). We included 165 cases in the study. Some patients were included in our previous reports on morpho-functional MRI (8, 37–39).

2.2 Magnetic resonance imaging

We performed standardized chest MRI after the initial CF diagnosis or after referral to our center as early as at the age of 3 months. We repeated exams annually using two 1.5 T scanner models from the same manufacturer (Magnetom Symphony and Magnetom Avanto, Siemens Healthcare, Erlangen, Germany). We kept the scanning protocol constant during the study period, apart from minor updates to new software versions as previously described (8–16, 24). We acquired T1-weighted sequences before and after intravenous application of contrast material and T2-weighted sequences before contrast. Children aged 5 years and younger were routinely sedated with oral or rectal chloral hydrate (100 mg/kg body weight, maximum dose of 2 g).

2.3 Staging CF lung disease

One observer (MOW) with more than 15 years of experience in chest MRI, who also evaluated all previous studies, assessed all MRI examinations using the established chest MRI scoring system (8–11, 13–15, 40). The MRI scoring system assigns a numerical disease severity score to each lobe (e.g., 0 = no presence, 1 = <50% of a lobe affected, and 2 = $\geq 50\%$ of a lobe affected) for each of the morphological score items bronchiectasis/wall thickening, mucus plugging, sacculation/abscess, consolidation, and special finding/pleural lesion, as well as for perfusion abnormalities. The sum of morphological findings becomes the MRI morphology score, perfusion abnormalities create the MRI perfusion score, and the sum of both results in the MRI global score, ranging from 0 to 72.

TABLE 1 MRI sequence details.

	BLADE	VIBE	HASTE
Slice thickness (mm)	4	4	6
Pixel spacing in plane (mm)	0.9375*0.9375	0.78125*0.78125	0.839*0.839
(min-max)	1.25*1.25	0.879*0.879	1.875*1.875
Matrix (min-max)	320*320	512*512	512*512
	384*384		

Since image acquisition protocols changed slightly over the years, pixel spacing and matrix size have different min/max values.

2.4 Image sequence selection

Three MRI sequences in coronal orientation were used (Table 1):

- 1 Balanced Steady State Free Precession Line Acquisition with Undersampling (BLADE): This is a T2-weighted turbo spin echo-based 2D sequence designed to reduce motion artifacts in MRI. It is particularly useful for imaging areas of the body that are prone to movement, like the lungs, or for imaging patients who have difficulty remaining still (41). Its acquisition can be split among multiple breath-holds (i.e., slices are not necessarily at the same depth of inspiration) or triggered using a navigator signal.
- 2 Volumetric Interpolated Breath-Hold Examination (VIBE): This is a T1-weighted 3D gradient echo sequence acquired after injection of a contrast agent. It was acquired in a single breath-hold and allows for high spatial resolution (42).
- 3 Half-Fourier Acquisition Single-Shot Turbo Spin-Echo (HASTE): This is a T2-weighted turbo spin echo 2D sequence that acquires each slice from a single echo train, minimizing motion effects at the cost of noticeable blurring in the phase encoding direction (43).

2.5 Dataset composition

From our database with 1,312 CF examinations from 266 patients, we selected 55 examinations for each MRI sequence (BLADE, VIBE, and HASTE), resulting in an overall 165 examinations from 84 patients (Figure 1). All cases were chosen to ensure an even distribution of age and gender, and to include varying levels of disease severity based on the global MRI score. To achieve this, the overall distribution of age, gender and disease severity was visualized and cases were then selected manually. From this overall dataset with 165 cases, 45 cases (15 for each sequence) were selected in a stratified manner, to represent the underlying distribution of age, gender, and global MRI score for the creation of the internal test set. This internal test set was not used for training, and solely utilized to test the final performance of the networks. In the internal test set, the median age was 9 years (± 4.92) (range 2 months–17 years) (Table 2, internal test set) with 46.7% male cases. The remaining cases were used for training the neural networks in the so-called training set. The training set had a median age of 9 years (± 4.78) (range 2 months–17 years) and 49.1% male cases (Table 3;

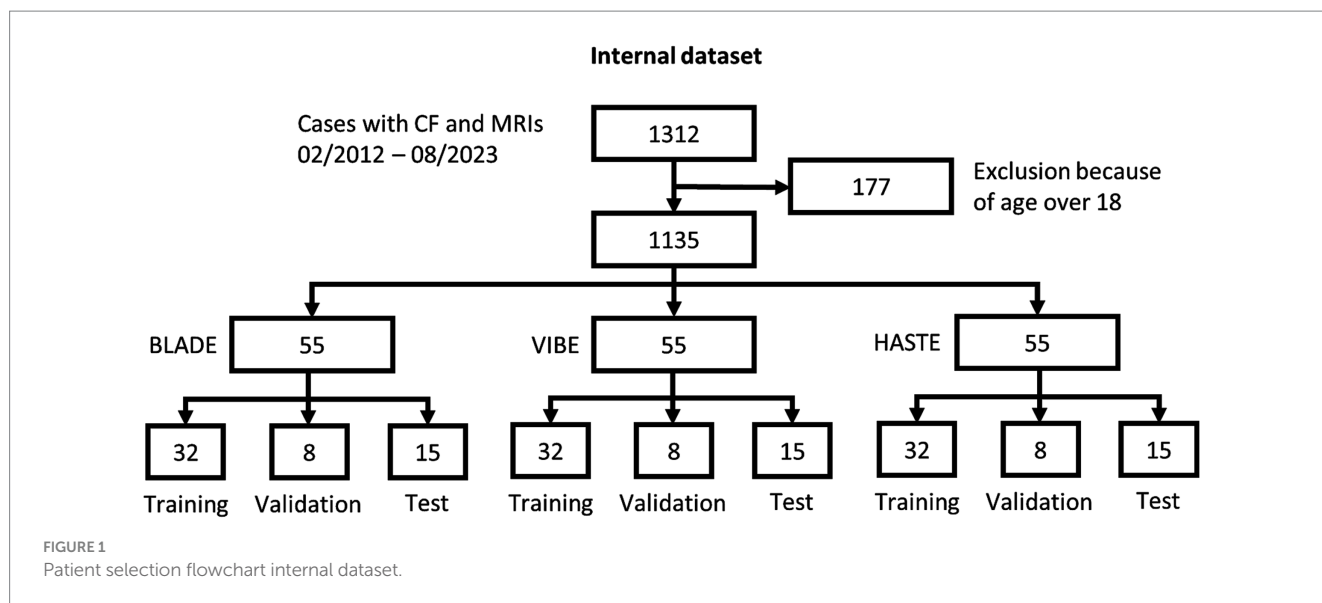


TABLE 2 Overview of the internal and external test sets.

	Internal test set	External test set
Cases, <i>n</i>	45	9
Age (years) median [range] (std)	9 [0.13–17.1] ± 4.92	0.79 [0.16–2.0] ± 0.63
Sex (m/f)	21 / 24	4/5
Height (cm) median [range] (std)	133.45 [93.0–174.8] ± 24.43	75.16 [62.0–83.0] ± 7.73
Weight (kg) median [range] (std)	28.03 [12.4–55.9] ± 14.65	9.2 [6.0–10.6] ± 1.73
CFTR genotype, <i>n</i> (%)	38 (84)	4 (44)
<i>F508del/F508del</i>	18 (40)	2 (22)
<i>F508del/other</i>	15 (33)	3 (33)
<i>Other/other</i>	4 (9)	
Pancreatic insufficiency, <i>n</i> (%)	31 (67)	9 (100)
Spirometry, <i>n</i> (%)	29 (65)	0 (0)
<i>ppFEV1</i> median [range] (std)	97.7 [62.2–123.7] ± 15.8	-
Multiple breath washout, <i>n</i> (%)	26 (58)	6 (67)
<i>LCI N₂</i> median [range] (std)	7.74 [4.74–10.57] ± 1.4	7.8 [6.58–9.05] ± 1.2
Global MRI Score median [range] (std)	12 [0–43] ± 9.8	10.22 [7–16] ± 2.72

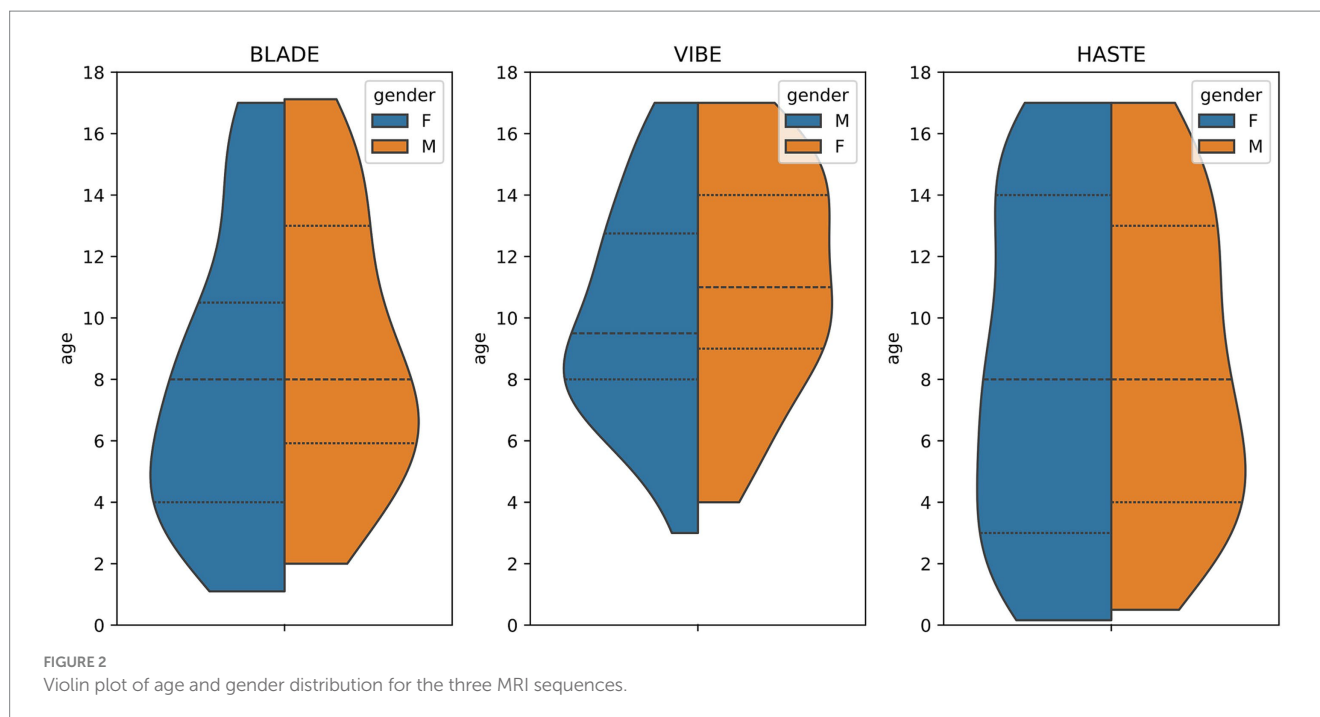
Figures 1, 2). Selecting the cases in such a way may allow good segmentations over all age and disease classes. We included cases only if relevant image data were available. No cases were excluded due to artifacts or poor quality. With regard to similarity, no notable differences were observed in the global MRI score across the three utilized sequences ($p=0.78$). A notable difference in age was observed between the three sequences ($p=0.006$). Patients undergoing imaging using HASTE were

TABLE 3 Patient characteristics of the internal dataset (training set).

	Internal training set
Cases, <i>n</i>	120
Age (years) median [range] (std)	9 [0.16–17.0] ± 4.78
Sex (m/f)	59/61
Height (cm) median [range] (std)	135 [52.7–175.3] ± 24.80
Weight (kg) median [range] (std)	28.2 [12.8–70.9] ± 16.05
CFTR genotype, <i>n</i> (%)	108 (90)
<i>F508del/F508del</i>	44 (37)
<i>F508del/other</i>	48 (40)
<i>Other/other</i>	12 (10)
Pancreatic insufficiency, <i>n</i> (%)	90 (75)
Spirometry, <i>n</i> (%)	46 (38)
<i>ppFEV1</i> median [range] (std)	91.58 [42.8–108.5] ± 16.3
Multiple breath washout, <i>n</i> (%)	45 (37.5)
<i>LCI N₂</i> median [range] (std)	8.28 [3.91–15.3] ± 2.46
Global MRI score median [range] (std)	12 [0–39] ± 7.83

notably younger, as HASTE is a contrast agent-free alternative to VIBE. The comparison between the internal training and test data revealed no statistically significant changes in either age ($p=0.06$) or global MRI score ($p=0.97$). The available data from all three sequences resulted in 6,010 2D slices. The training set comprised a total of 4,290 slices, with an average of 34 slices for BLADE, 49 slices for VIBE and 24 slices for HASTE per MRI. In the internal test set, a total of 1,720 slices were used, with an average of 36 slices for BLADE, 52 slices for VIBE, and 27 slices for HASTE.

Additionally, we collected an external dataset from two different centers (Center A: one case, Center B: eight cases) comprising nine HASTE acquisitions from nine cases (Table 2) (44). Compared to the internal dataset, the age distribution of the external test set was dominated by very young patients (Table 2). This was reflected in the statistically significant difference in age between the internal and



external datasets ($p < 0.0005$). Regarding disease severity, however, the external patients were similar to those of the internal dataset. A comparison of the global MRI score exhibited no notable differences ($p = 0.86$). The external test set had an overall number of 174 2D slices with an average of 19 slices per MRI.

2.6 Segmentation ground truth generation

We manually segmented all MRIs using three independent observers with 1 (JM), 2.5 (FGR), and 5 years (LW) experience in lung MRI segmentation, respectively. They created reference segmentations of the lung halves using the open-source software Medical Imaging Interaction Toolkit (MITK, version 2021.10) in combination with a Wacom Cintiq 16 tablet and pen. In the event of disagreement among observers, agreement was reached by individual comparison, collective discussion and consensus among the observers.

2.7 Segmentation questionnaire

In cooperation with the experienced radiologists, we designed a qualitative questionnaire for fine-grained evaluation of the segmentations (Supplementary Figure 1). The questionnaire evaluated the overall segmentation quality on an 11-point Likert-scale (45), ranging from 0 (worst quality) to 10 (best quality). Furthermore, the questionnaire included a detailed evaluation of the lung segmentation and specific information regarding the segmentation performance in specific anatomical regions (ventrally, dorsally, mediastinum, periphery, apex, and diaphragm). Information on incomplete segmentation or over-segmentation in specific areas could be provided. Lastly, the observer was given the opportunity to provide an open text response to the segmentation.

2.8 Slice based qualitative analysis

To gain further insight into the segmentation quality, all lungs from the internal test set were subjected to a detailed examination by a radiologist to identify any instances of incorrect segmentation. For each lung, the number of slices requiring correction was annotated. In conjunction with the data on the overall slices, this provides an indication of the quantity of usable slices. The number of slices requiring correction is reported as a mean percentage, with standard deviation and maximum.

2.9 nnU-Net implementation

The latest implementation of the 2D nnU-Net (Version 2) was utilized in its default configuration. It is a self-configuring framework, which automatically adapts its architecture, pre-processing, and training pipeline to a given dataset. The nnU-Net framework employs a U-Net-based architecture comprising an encoder-decoder structure. On the encoder path, the spatial dimensions of the input image are successively reduced through convolutional layers and max-pooling, thereby capturing increasingly abstract feature representations. On the decoder path, upsampling is applied to restore the spatial dimensions, concatenating feature maps from the corresponding encoder layers. This allows for high-level semantic information and precise localization. For further details to the nnU-Net, please refer to (22). Three individual nnU-Net configurations were trained, one for each sequence using the following steps: based on the 55 study cases per sequence, the data were partitioned into 58% as training set, 14% as validation set, and 27% as test set. This resulted in 32 cases being used for training, eight cases for validation and 15 cases for testing per sequence. The training and validation sets were utilized for the initial and fine-tuning training of the neural

network, while the test set was withheld for final evaluation. To ensure an average performance, we validated the models utilizing 5-fold cross-validation with different training and validation set partitions as per default nnU-Net configuration. All calculations were performed on two Tesla V100S PCIe 32GB with 1,000 epochs and an average run time of 33 h per fold. The batch size was subject to variation during the training phase, with values of 14, 32, and 33, respectively, being applied to BLADE, VIBE, and HASTE. Stochastic gradient descent was employed for optimization purposes, with a weight decay of $3e^{-5}$ and an initial learning rate of 0.01. Z-score normalization was utilized as the normalization method. For inference and producing the final predictions the nnU-Net uses an ensemble of all five folds, reporting one final result for the test set. For external validation a separate dataset was utilized. This external dataset was chosen to simulate real-world scenarios and challenges, ensuring a comprehensive examination of the model's performance across diverse imaging conditions.

2.10 Statistical analyses

A one-way ANOVA test was used to determine if there were statistically significant differences in global MRI score and ages among the different datasets. Results were considered significant at $p < 0.05$.

Furthermore, the Sørensen-Dice-Score (DSC), calculated from the spatial overlap between the ground truth segmentation (GT) and predicted segmentation (PS), was utilized to evaluate the entire MRI sequence (46). The DSC ranges from 0 to 1, evaluating the quality of the segmentation indicated by the overlap and is defined as follows:

$$DSC = \frac{2|PS \cap GT|}{|PS| + |GT|}$$

First, the Dice coefficient was calculated between each manual segmentation and predicted mask, and subsequently, the mean value was obtained for the entire stack of slices. This process was conducted for both the right and left lungs, as well as for the combination of both lung halves.

Data were analyzed with Python (Version 3.9) using the package SciPy (Version 1.11.4) (47). The Pearson correlation coefficient, indicating strength of linear relationship, was calculated for the DSC vs. age and DSC vs. the global MRI score (48). In general, the Pearson correlation coefficient measures the linear correlation of two sets of data and is defined as:

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}$$

Since the Sørensen-Dice-Score does not provide any indications regarding the location of incorrect segmentations or crucial errors, we deployed an additional questionnaire, which was filled out once for each internal case. To assess the generalizability and robustness of lung segmentation, we conducted an evaluation using an external dataset distinct from the training and validation sets. Due to data availability, only the HASTE model was tested.

3 Results

3.1 Internal and external test set demographics

A total of 45 cases were utilized for the internal test set, while the external test set consisted of nine cases from two distinct centers. The cases from the external dataset are notably younger, with a median age of 0.79 years, whereas the internal test set had a median age of 9 years (Table 2). With regard to the global MRI score, the external dataset exhibited a slightly lower median of 10.22, as compared to the internal test set, which had a median global MRI score of 12.

3.2 BLADE, HASTE, and VIBE are equally well suited for nnU-Net training

Using VIBE and BLADE, the nnU-Net achieved a mean DSC of 0.96 (Table 4; Figures 3, 4). HASTE demonstrated comparable performance with a mean DSC of 0.95. For the BLADE sequence, the right lung exhibited slightly superior segmentation, whereas both lungs demonstrated equivalent performance in the VIBE sequence. On the HASTE sequence, the left lung reached a higher DSC compared to the right lung with a DSC of 0.96 and 0.93, respectively (Table 4).

3.3 Questionnaire confirms segmentation quality

Our analysis of the questionnaire for the 45 internal test cases showed similar results to the overall high Dice coefficients. The segmentations derived from all three sequences were evaluated with a median score of nine out of 10 points (9/10) on the Likert Scale, with a standard deviation of 2.02, 1.48, and 1.18 for BLADE, VIBE, and HASTE, respectively (Figure 5). Additional information can be found in Supplementary Figures 2, 3. In addition to the quality of the segmentation, the observer provided information about inconsistencies or errors in the segmentations. Three general trends were identified (Supplementary Figure 3):

1. missing ventral segmentations;
2. missing segmentations near the costodiaphragmatic recess; and
3. incorrect segmentation of the lower mediastinum.

Further, in some cases, segmentations were incomplete in the lung periphery, leaving a small space unaccounted for close to the edge of the lung (Figure 6).

TABLE 4 Sørensen-Dice-Score (DSC) results for three sequences, showing mean (stdv).

DSC	BLADE	VIBE	HASTE
Whole lung	0.96 ± 0.05	0.96 ± 0.04	0.95 ± 0.05
Left lung	0.95 ± 0.09	0.96 ± 0.04	0.96 ± 0.03
Right lung	0.97 ± 0.03	0.96 ± 0.04	0.93 ± 0.10

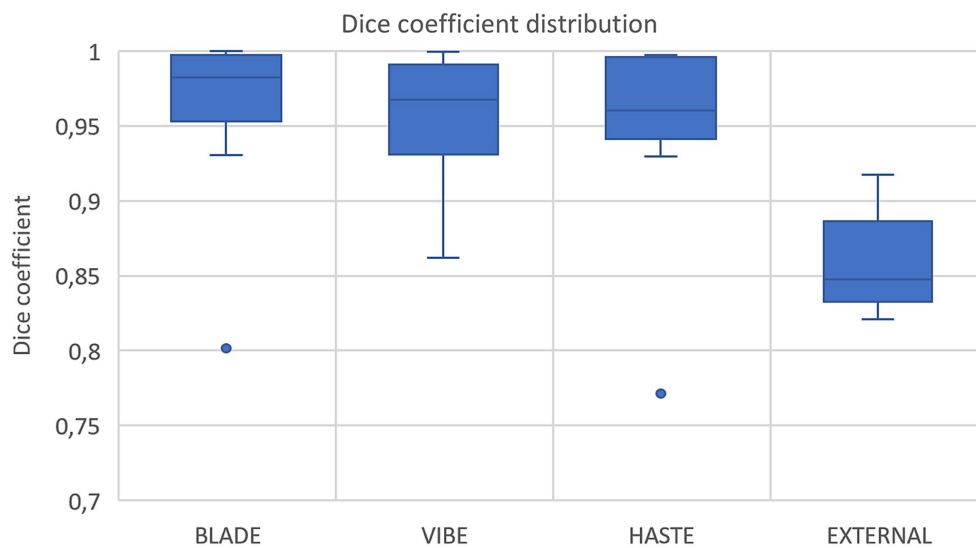


FIGURE 3
Box plots with dice coefficient of internal test sets from BLADE, VIBE, and HASTE and the external test set.

3.4 nnU-Net performance is independent of age and disease severity

Anatomy such as form and size of the chest change with age, and lung disease severity alters anatomy and signal of the lungs. Thus, we correlated DSC with the patient age and DSC with disease severity. Age as well as the MRI global score did not show an association with nnU-Net performance $r=0.09$ and $r=-0.12$, respectively (Supplementary Figures 4, 5).

3.5 nnU-Net shows acceptable performance on external validation data

Nine cases with corresponding HASTE MRI from two external centers, with ages ranging from 3 months to 2 years were segmented using the network trained on the HASTE imaging data. The average DSC across all validated cases was $0.85 (\pm 0.03)$ with a range from 0.82 to 0.92 for both lung halves combined (Figure 3, external data). Regarding lung halves, the left lung was segmented better, with a DSC ranging from 0.79 to 0.92, compared to the right lung with a DSC of 0.70 to 0.92. This indicates acceptable, but not perfect performance.

3.6 Slice based analysis highlights segmentation quality

A visual inspection was conducted on all data from the internal test set to ascertain the quantity of slices that would require manual correction. A total of 1,720 slices from the 45 internal test cases were subjected to quality control. The results are consistent with the responses provided in question 1 of the questionnaire. The CF case with the lowest score assigned by the radiologist (2/10) exhibited the highest number of slices requiring correction. Specifically, 79% of slices in the right lung and 29% of slices in the left lung were of

insufficient quality. Overall, the mean percentage of slices in the right lung and left lung that required correction was 10.60% (± 16.46) and 8.75% (± 9.39), respectively (Supplementary Figure 6).

4 Discussion

Segmentation can play a vital role as a pre-processing step before applying machine learning-based image analysis methods. In our work, lung half segmentation of pediatric MRIs of CF patients using three different sequences, BLADE, VIBE, and HASTE were created utilizing the nnU-Net neural network. A dataset comprising 165 cases, with 55 cases for each of the three sequences, was employed for the training, validation, and testing of the nnU-Net. For each sequence, the nnU-Net was trained individually using a training set of 40 cases and a testing set of 15 cases. For evaluation, the Sørensen-Dice-Score was used in combination with a tailored questionnaire and a slice-based analysis to provide a more detailed insight into the quality of the segmentations.

Overall, the segmentation performance achieved a mean Dice of 0.95 or higher for all sequences and lung halves except for the right lung on the HASTE sequence, which reached a mean Dice of 0.93. With the patient's age ranging from just a few months to 17 years, the segmentation performance was correlated with age. Generally, it was visible that the segmentation quality stayed constant across all pediatric age classes, further supported by Pearson correlation coefficient $r=0.09$. Due to the different disease status of the patients, the global MRI score was correlated with the Dice coefficient. Patients with both lower and higher global MRI score were segmented equally well, which is supported by the Pearson correlation coefficient of $r=-0.12$. This demonstrates the excellent performance of the nnU-Net for lung lobe segmentation in pediatric chest MRIs within our cohort. The high mean DSC indicates robust segmentation performance, independent of the underlying pathological changes induced by CF in the pediatric stage. An improvement in segmentation

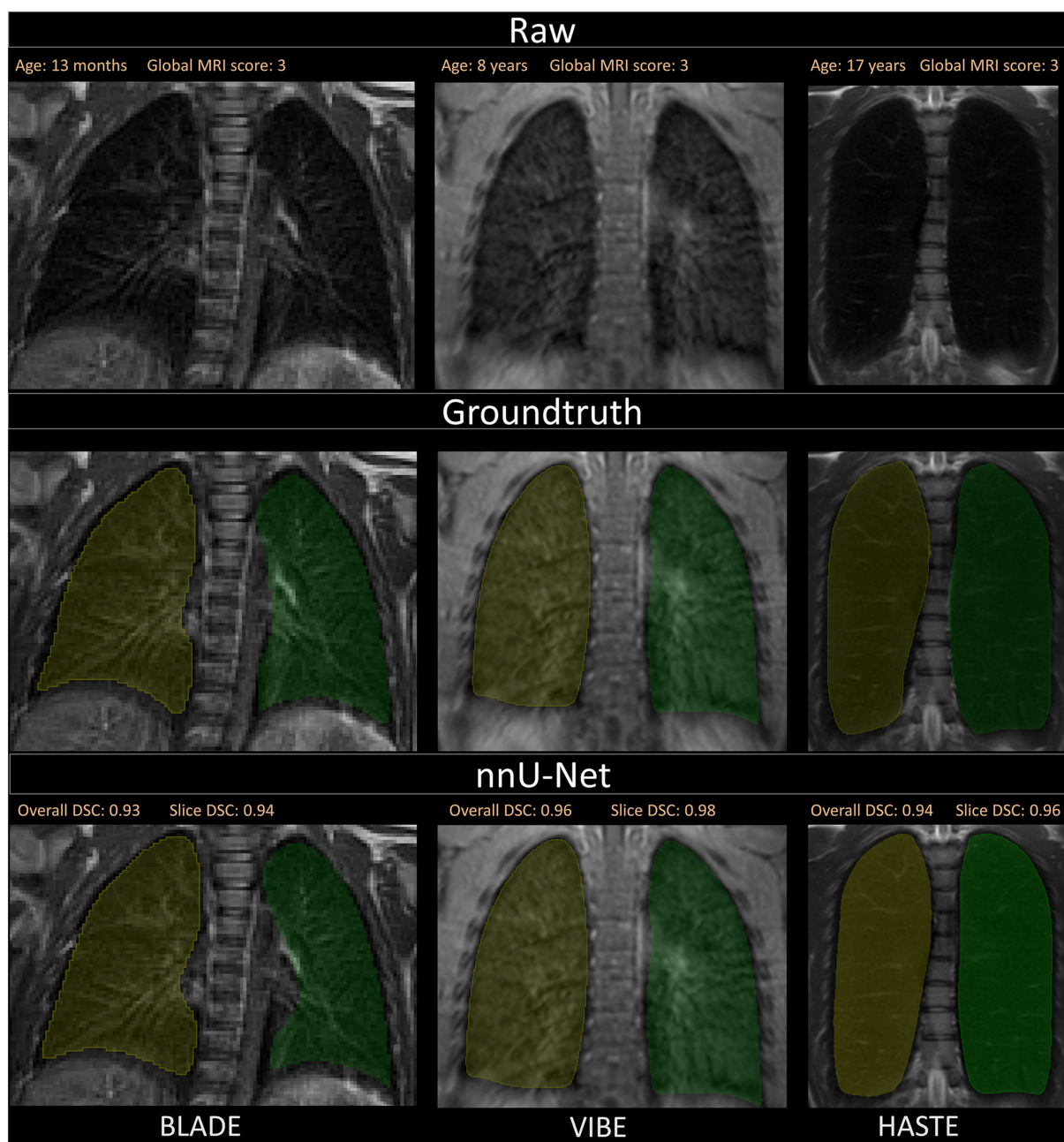


FIGURE 4

Visualization of the three different sequences with ground truth segmentations and good segmentations produced by the nnU-Net. Overall DSC corresponds to the dice coefficient of the entire lung and Slice DSC to the dice coefficient of the visualized slice. The segmentation of the right lung is indicated in yellow and the left lung segmentation in green. Each column corresponds to one MRI sequence. In the top row, the raw images are shown. The second row contains the manually annotated lung halves (ground truth). In the third row, the segmentation calculated by the corresponding nnU-Net is depicted. The three shown patients are of ascending age from left to right, thus the different lung sizes. All three patients have a global MRI score of 3. The results of the questionnaire indicated that the lungs were rated with a score of 10/10 for BLADE, 9/10 for VIBE, and 9/10 for HASTE. The different contrasts and gray levels are due to the different sequences. Both the ground truth and the segmentation of the lung halves appear to be very similar. Although the right and left lung differ in size and shape, segmentation performance seems to be almost equal. In general, the high Sørensen-Dice-Score and corresponding high segmentation performance are evident.

performance might be expected with an increased amount of training data (49). However, in segmentation tasks that require precise ground truth annotations, which are extremely time-intensive to generate, necessary trade-offs must be made.

For a qualitative analysis, we provided a questionnaire to the observers for the purpose of evaluating the segmentations manually

in addition to the Sørensen-Dice-Score. Consistent segmentation errors in the ventral and dorsal areas of the lung, as well as around the costodiaphragmatic recess were detected. These errors can be caused by the thickness of the image slices, which directly affects the appearance of the tissue. When the slice thickness increases, tissue other than lung tissue becomes included, which may lead to the partial



volume effect (PVE) (50). PVE occurs in volumetric imaging, when more than one tissue type is present in a voxel. In cases where the lung parenchyma ends in the middle of the slice, the voxel will have a different shade of gray compared to voxels completely inside or outside the lung. Depending on the patient and amount of non-lung area on the entire slice, this gray level complicates manual and automated segmentation (Figure 7). Furthermore, even the slightest movement in the ventral and dorsal areas may introduce additional artifacts or blurring. While the observers annotated certain areas as lung tissue, the neural network failed to do so. To overcome this challenge, more annotated data could improve the dorsal and ventral segmentation performance. In addition to the questionnaire, the observer conducted a slice-based analysis annotating which slices required manual correction. The analysis revealed that the left lung necessitated more corrections than the right lung.

To achieve a more general evaluation of the trained neural networks, an external dataset from two other centers was segmented and evaluated with the Sørensen-Dice-Score. This was a useful test to explore whether generalization had been achieved, allowing the processing of data from a different source than the training data. Generally, overfitting on the training data is common, leading to very good performances on the training and test set from the same distribution but poor performance on external data. Despite the fact that the Sørensen-Dice-Score for the external dataset did not exceed 0.92, with a range of 0.7–0.92, it suggests that the trained neural network has overall generalizability, given the differences between the two test sets regarding age and number of slices. A comparison of the number of MRI slices in the internal and external test sets reveals notable differences: the internal dataset averages 38 slices per MRI, while the external dataset averages only 19 slices. Since the Dice coefficient is more sensitive to segmentation errors when the overall segmented area is smaller, this difference in slice count and corresponding segmentation area must be taken into account when interpreting the results (51).

Factors such as different MRI scanner specifications, protocols, slice thickness, and resolution affect image quality and therefore

segmentation performance. The small number of external MRIs ($n=9$) limited the general interpretability. Efforts to improve segmentation performance on the external dataset could include retraining the nnU-Net configuration with external MRI data to reduce segmentation errors. The overall segmentation performances on the three MRI sequences were comparable with existing work in literature. Lung CT scans have been segmented fully automated, reaching high accuracies for lung lobe segmentations with a mean Dice coefficient of up to 0.97 (52). In chest MRI, lung segmentation can be achieved using traditional approaches such as thresholding, but neural network-based segmentation approaches have recently been shown to outperform traditional methods (25). Astley et al. even showed the nnU-Net can be trained to perform well across several sequences, diseases, and vendors reaching a median Dice coefficient of 0.96 on the internal and 0.97 on an independent test set (30). Moreover, their results demonstrated that the 3D-Unet exhibited superior performance compared to the 2D version, which, in turn, outperformed the conventional segmentation approach, spatial fuzzy C-means. In contrast to their work, our study focused on pediatric MRIs of the entire pediatric range of patients with varying degrees of CF disease severity. Efforts toward improving MRI-based lung segmentation include artificially created images to increase robustness in case of severe pathologies (31). For hyperpolarized ^{129}Xe MRI, segmentation performances with a Dice score of 0.929 and above were demonstrated using multiple different methods, highlighting the superiority of the nnU-Net over conventional segmentation methods (53, 54). Neonatal lung segmentations showed a Sørensen-Dice-Score of 0.908 and 0.880 on an independent test set with segmentations automatically by a combination of U-Nets (32).

The resulting segmentations for both the internal and external test sets of the underlying study exhibited variability, yet never attained a Dice coefficient of 1.0. This raises a pivotal question about the criteria for determining whether a segmentation is suitable for subsequent processing or downstream analyses. While a Dice coefficient of 1.0 represents perfect segmentation, striving to improve the coefficient from an already high mean value such as 0.95, may demand a

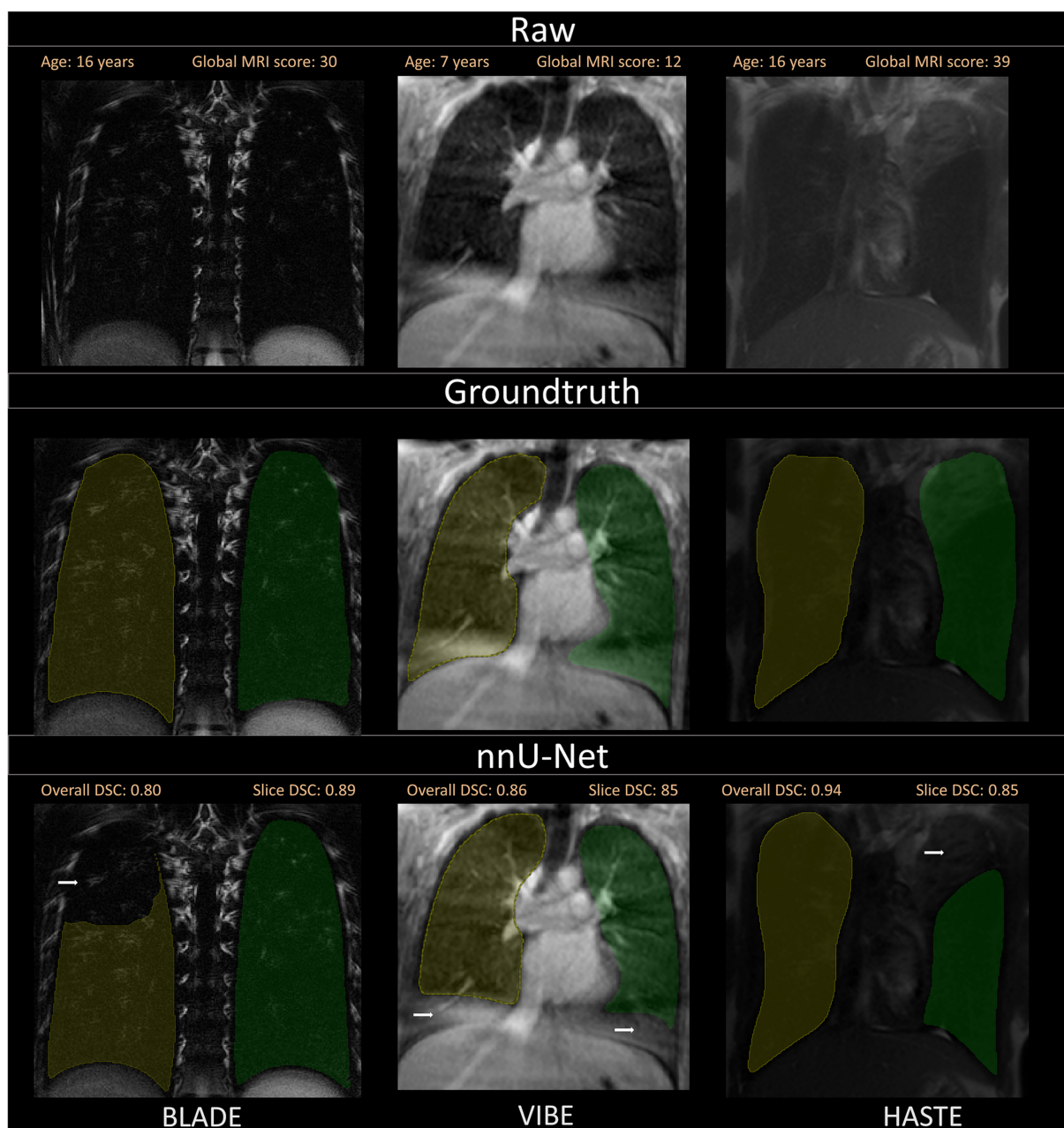


FIGURE 6

A selection of segmentations with a slightly lower Sørensen-Dice-Score, as well as visual discrepancies between ground truth and nnU-Net segmentation is shown. Overall DSC corresponds to the dice coefficient of the entire lung and Slice DSC to the dice coefficient of the visualized slice. Segmentation errors are indicated with white arrows in the second row. Three common segmentation mistakes are shown: Incomplete segmentations for BLADE, wrong segmentations due to breathing motion or other artifacts for VIBE and pathological changes influencing segmentation performance on the patient captured with the HASTE protocol. Based on the results of the questionnaire, the lungs were rated with a score of 2/10 for BLADE, 6/10 for VIBE, and 7/10 for HASTE.

disproportionate amount of time, effort and computational resources. In practice, the pursuit of marginal improvements—bringing the Dice score closer to 1.0—often results in diminishing returns. Such refinements may have minimal impact on the overall effectiveness or accuracy of downstream tasks, particularly when the current segmentation quality is already deemed suitable for clinical decision-making or research purposes. Therefore, it is critical to assess whether the additional time and computational effort invested in further optimizing segmentation is justified, or whether the existing

performance is sufficient for the intended applications. When observers segment lungs, they hardly ever reach complete agreement. Segmentation tasks are always dependent on the reader, their experience in the domain, and the tools used. Literature has shown that a Dice coefficient below 0.9 is not uncommon as reader agreement (55). Given that the overarching objective is to automate the Eichinger score, it can be argued that segmentation errors that do not significantly impact the majority of a lung half might be considered acceptable. However, in other research questions, this threshold may

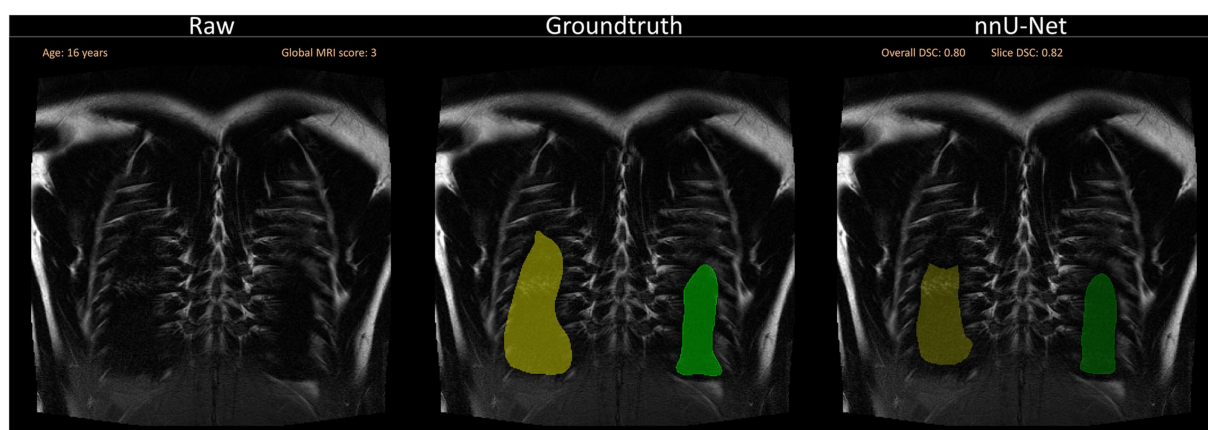


FIGURE 7

Segmentation errors dorsally due to partial volume effects and general difficulty to segment close to the ribs. The visualization shows a slice close to the back of the patient without any segmentations (left), with the annotated ground truth (middle) and the suggested segmentation by the neural network (right). Overall DSC corresponds to the dice coefficient of the entire lung and Slice DSC to the dice coefficient of the visualized slice.

have to be set differently, for example, in the context of tumor resection or radiotherapy planning, where segmentations require a higher degree of accuracy (56, 57).

In summary, the results obtained in this study are comparable to those reported in similar studies. To the best of our knowledge, we are the first to demonstrate successful pediatric lung half segmentations for patients with different stages of cystic fibrosis on the MRI sequences BLADE, VIBE, and HASTE.

Our study has some limitations that require discussion. The sample size of 55 cases with corresponding MRIs for each sequence, especially in the age group of patients under 1 year, is relatively small. Compared to the external dataset, all internal cases have larger lungs due to the higher mean age, which could influence the segmentation performance. Moreover, the majority of included patients had a global MRI score of 20 or less. Therefore, it is unclear whether our results are transferable to cohorts of older patients or patients with more advanced lung disease. Future work may focus on this aspect as well as an extension to other sequences.

Recent advancements in this field of research are driving the development of various methods for automated segmentation (19). In the future, it may be valuable to explore these approaches on this dataset and consider expanding the current model to include the remainder of the patients and cases. Especially, since the overall goal of automating the Eichinger score, works toward automated lung lobe segmentations should be explored. Pusterla et al. showed recently that automated lung segmentation with a combination of neural networks is possible with high accuracy (35). Earlier studies demonstrated that segmentation of perfusion maps with a 3D U-Net is an effective approach. However, the evaluation of lung lobes on MRI is challenging due to the difficulty in discerning lobe fissures, if they are visible at all. A lung atlas-based approach, which is independent of age and disease status, may prove advantageous, particularly in light of the findings reported by Tutison et al. regarding the segmentation of the lung (58). With the automated lung segmentation in place, further complex deep learning-based analysis techniques can be applied to assist radiologists in monitoring treatment response, therapy progression, and overall lung health of CF patients, potentially saving time. These results reinforce the efforts

toward automated analysis of chest MRIs of patients with cystic fibrosis.

In conclusion, the performance of the nnU-Net in segmenting the lung halves of MRIs from pediatric CF patients demonstrated good agreement with manual segmentations. The segmentation performance of pediatric CF patients does not appear to be significantly influenced by age or disease status.

Data availability statement

The data analyzed in this study are subject to the following licenses/restrictions: the data that support the findings of this work are available from the corresponding author upon reasonable request. Requests to access these datasets should be directed to urs.eisenmann@med.uni-heidelberg.de.

Ethics statement

The studies involving humans were approved by clinicaltrials.gov identifiers NCT00760071 and NCT02270476. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation in this study was provided by the participants' legal guardians/next of kin.

Author contributions

FR: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. LW: Writing – review & editing, Visualization, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. NH: Writing – review & editing, Visualization, Validation, Supervision, Resources, Project administration, Methodology,

Investigation, Formal analysis, Data curation, Conceptualization. JM: Writing – review & editing, Validation, Software, Methodology, Investigation, Data curation. SK: Writing – review & editing, Visualization, Validation, Software, Resources, Methodology, Investigation, Data curation. ME: Writing – review & editing, Visualization, Resources, Project administration, Methodology, Formal analysis, Data curation. MS: Writing – review & editing, Validation, Resources, Investigation, Data curation. ST: Validation, Writing – review & editing, Resources, Methodology, Data curation. PL-S: Writing – review & editing, Resources, Data curation. SoG: Writing – review & editing, Resources, Data curation. SiG: Writing – review & editing, Resources, Data curation. H-UK: Writing – review & editing, Resources, Funding acquisition. AA: Writing – review & editing, Resources, Data curation. J-PS: Writing – review & editing, Resources, Data curation. OS: Writing – review & editing, Resources, Data curation. MM: Writing – review & editing, Resources, Data curation. PK: Writing – review & editing, Validation, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Data curation. MW: Conceptualization, Writing – review & editing, Visualization, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation. UE: Writing – review & editing, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This work was supported by the Vertex Innovation Award 2022, grants from the German Federal Ministry of Education and Research (82DZL00401, 82DZL004A1, and 82DZL009B1), the German Research Foundation (STA 1685/1-1), and the Mukoviszidose e.V. (S02/09, C-H-P 1504). SiG and MS are participants of the BIH-Charité Clinician Scientist Program funded by the Charité-Universitätsmedizin Berlin and the BIH. Funders were not involved in the collection, analysis and

References

- Gibson RL, Burns JL, Ramsey BW. Pathophysiology and Management of Pulmonary Infections in cystic fibrosis. *Am J Respir Crit Care Med.* (2003) 168:918–51. doi: 10.1164/rccm.200304-505SO
- Welsh MJ, Ramsey BW, Accurso F, Cutting GR. Cystic fibrosis In: DL Valle, S Antonarakis, A Ballabio, AL Beaudet and GA Mitchell, editors. *The Online Metabolic and Molecular Bases of Inherited Disease.* New York, NY: McGraw-Hill Education (2019)
- Bell SC, Mall MA, Gutierrez H, Macek M, Madge S, Davies JC, et al. The future of cystic fibrosis care: a global perspective. *Lancet Respir Med.* (2020) 8:65–124. doi: 10.1016/S2213-2600(19)30337-6
- Grasemann H, Ratjen F. Early lung disease in cystic fibrosis. *Lancet Respir Med.* (2013) 1:148–57. doi: 10.1016/S2213-2600(13)70026-2
- Stick S, Tiddens H, Aurora P, Gustafsson P, Ranganathan S, Robinson P, et al. Early intervention studies in infants and preschool children with cystic fibrosis: are we ready? *Eur Respir J.* (2013) 42:527–38. doi: 10.1183/09031936.00108212
- Sly PD, Wainwright CE. Diagnosis and early life risk factors for bronchiectasis in cystic fibrosis: a review. *Expert Rev Respir Med.* (2016) 10:1003–10. doi: 10.1080/17476348.2016.1204915
- Wielpütz MO, Eichinger M, Biederer J, Wege S, Stahl M, Sommerburg O, et al. Bildgebung der Lunge bei Mukoviszidose und klinische Interpretation. *RöFo.* (2016) 188:834–45. doi: 10.1055/s-0042-104936

interpretation of data, in the writing of the report and in the decision to submit the article for publication.

Acknowledgments

This work was supported by the Vertex Innovation Award 2022 by Vertex Pharmaceuticals. For the publication fee we acknowledge financial support by Heidelberg University. Parts of this manuscript were improved for readability purposes with AI (DeepL Write).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer OP declared a past co-authorship with the authors MW and MM to the handling editor.

The author(s) declared that they were an editorial board member of *Frontiers*, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmed.2024.1401473/full#supplementary-material>

- Wielpütz MO, Puderbach M, Kopp-Schneider A, Stahl M, Fritzsche E, Sommerburg O, 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) 189:956–65. doi: 10.1164/rccm.201309-1659OC
- Stahl M, Wielpütz MO, Graeber SY, Joachim C, Sommerburg O, Kauczor H-U, et al. Comparison of lung clearance index and magnetic resonance imaging for assessment of lung disease in children with cystic fibrosis. *Am J Respir Crit Care Med.* (2017) 195:349–59. doi: 10.1164/rccm.201604-0893OC
- Wielpütz MO, von Stackelberg O, Stahl M, Jobst BJ, Eichinger M, Puderbach MU, et al. Multicentre standardisation of chest MRI as radiation-free outcome measure of lung disease in young children with cystic fibrosis. *J Cyst Fibros.* (2018) 17:518–27. doi: 10.1016/j.jcf.2018.05.003
- Wielpütz MO, Eichinger M, Wege S, Eberhardt R, Mall MA, Kauczor H-U, et al. Midterm reproducibility of chest magnetic resonance imaging in adults with clinically stable cystic fibrosis and chronic obstructive pulmonary disease. *Am J Respir Crit Care Med.* (2019) 200:103–7. doi: 10.1164/rccm.201812-2356LE
- Woods JC, Wild JM, Wielpütz MO, Clancy JP, Hatabu H, Kauczor H-U, et al. Current state of the art MRI for the longitudinal assessment of cystic fibrosis. *J Magn Reson Imaging.* (2020) 52:1306–20. doi: 10.1002/jmri.27030
- Stahl M, Steinke E, Graeber SY, Joachim C, Seitz C, Kauczor H-U, et al. Magnetic resonance imaging detects progression of lung disease and impact of newborn screening

- in preschool children with cystic fibrosis. *Am J Respir Crit Care Med.* (2021) 204:943–53. doi: 10.1164/rccm.202102-0278OC
14. Wucherpfennig L, Triphan SMF, Wege S, Kauczor H-U, Heussel CP, Sommerburg O, et al. Elexacaftor/Tezacaftor/Ivacaftor improves bronchial artery dilatation detected by magnetic resonance imaging in patients with cystic fibrosis. *Ann Am Thorac Soc.* (2023) 20:1595–604. doi: 10.1513/AnnalsATS.202302-168OC
15. Eichinger M, Optazait D-E, Kopp-Schneider A, Hintze C, Biederer J, Niemann A, et al. Morphologic and functional scoring of cystic fibrosis lung disease using MRI. *Eur J Radiol.* (2012) 81:1321–9. doi: 10.1016/j.ejrad.2011.02.045
16. Puderbach M, Eichinger M, Haeselbarth J, Ley S, Kopp-Schneider A, Tuengerthal S, et al. Assessment of morphological MRI for pulmonary changes in cystic fibrosis (CF) patients. *Investig Radiol.* (2007) 42:715–24. doi: 10.1097/RLI.0b013e318074fd81
17. Asgari Taghanaki S, Abhishek K, Cohen JP, Cohen-Adad J, Hamarneh G. Deep semantic segmentation of natural and medical images: a review. *Artif Intell Rev.* (2021) 54:137–78. doi: 10.1007/s10462-020-09854-1
18. Wang R, Lei T, Cui R, Zhang B, Meng H, Nandi AK. Medical image segmentation using deep learning: a survey. *IET Image Process.* (2022) 16:1243–67. doi: 10.1049/ipr2.12419
19. Liu X, Song L, Liu S, Zhang Y. A review of deep-learning-based medical image segmentation methods. *Sustain For.* (2021) 13:1224. doi: 10.3390/su13031224
20. Kumar Y, Brar TPS, Kaur C, Singh C. A comprehensive study of deep learning methods for kidney tumor, cyst, and stone diagnostics and detection using CT images. *Arch Computat Methods Eng.* (2024) 31:4163–4188. doi: 10.1007/s11831-024-10112-8
21. Qureshi I, Yan J, Abbas Q, Shaheed K, Riaz AB, Wahid A, et al. Medical image segmentation using deep semantic-based methods: a review of techniques, applications and emerging trends. *Inform Fusion.* (2023) 90:316–52. doi: 10.1016/j.inffus.2022.09.031
22. Isensee F, Petersen J, Klein A, Zimmerer D, Jaeger PF, Kohl S, et al. (2018). nn U-Net: Self-adapting framework for U-net-based medical image segmentation. Available online at: <https://arxiv.org/pdf/1809.10486.pdf>
23. Siddique N, Paheding S, Elkin CP, Devabhaktuni V. U-net and its variants for medical image segmentation: a review of theory and applications. *IEEE Access.* (2021) 9:82031–57. doi: 10.1109/ACCESS.2021.3086020
24. Thukral BB. Problems and preferences in pediatric imaging. *Indian J Radiol Imag.* (2015) 25:359–64. doi: 10.4103/0971-3026.169466
25. Kohlmann P, Strehlow J, Jobst B, Krass S, Kuhnigk J-M, Anjorin A, 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) 10:403–17. doi: 10.1007/s11548-014-1090-0
26. Heimann T, Eichinger M, Bauman G, Bischoff A, Puderbach M, Meinzer H-P (2012). “Automated scoring of regional lung perfusion in children from contrast enhanced 3D MRI” in *Medical Imaging 2012: Computer-Aided Diagnosis*. SPIE. 83150U (SPIE Proceedings).
27. Zha W, Fain SB, Schiebler ML, Evans MD, Nagle SK, Liu F. Deep convolutional neural networks with multiplane consensus labeling for lung function quantification using UTE proton MRI. *J Magn Reson Imaging.* (2019) 50:1169–81. doi: 10.1002/jmri.26734
28. Weng AM, Heidenreich JF, Metz C, Veldhoen S, Bley TA, Wech T. Deep learning-based segmentation of the lung in MR-images acquired by a stack-of-spirals trajectory at ultra-short echo-times. *BMC Med Imaging.* (2021) 21:79. doi: 10.1186/s12880-021-00608-1
29. Willers C, Bauman G, Andermatt S, Santini F, Sandkühler R, Ramsey KA, et al. The impact of segmentation on whole-lung functional MRI quantification: repeatability and reproducibility from multiple human observers and an artificial neural network. *Magn Reson Med.* (2021) 85:1079–92. doi: 10.1002/mrm.28476
30. Astley JR, Biancardi AM, Hughes PJC, Marshall H, Collier GJ, Chan H-F, et al. Implementable deep learning for multi-sequence proton MRI lung segmentation: a multi-center, multi-vendor, and multi-disease study. *J Magn Reson Imaging.* (2023) 58:1030–44. doi: 10.1002/jmri.28643
31. Crisosto C, Voskrebenezv A, Gutberlet M, Klimes F, Kaireit TF, Pöhler G, et al. Artificially-generated consolidations and balanced augmentation increase performance of U-net for lung parenchyma segmentation on MR images. *PLoS One.* (2023) 18:e0285378. doi: 10.1371/journal.pone.0285378
32. Mairhormann B, Castelblanco A, Häfner F, Koliogiannis V, Haist L, Winter D, et al. Automated MRI lung segmentation and 3D morphologic features for quantification of neonatal lung disease. *Radiol Artif Intellig.* (2023) 5:e220239. doi: 10.1148/ryai.220239
33. Missimer JH, Emert F, Lomax AJ, Weber DC. Automatic lung segmentation of magnetic resonance images: a new approach applied to healthy volunteers undergoing enhanced deep-inspiration-breath-hold for motion-mitigated 4D proton therapy of lung tumors. *Phys Imag Radiat Oncol.* (2024) 29:100531. doi: 10.1016/j.phro.2024.100531
34. Taran TV, Pavlova OS, Gulyaev MV, Dmitriev DS, Pistrak AG, Ryabikov KN, et al. Automated image registration and perfusion sorting algorithms for PREFUL MRI. *Appl Magn Reson.* (2024) 55:741–52. doi: 10.1007/s00723-024-01684-6
35. Pusterla O, Heule R, Santini F, Weikert T, Willers C, Andermatt S, et al. MRI lung lobe segmentation in pediatric cystic fibrosis patients using a recurrent neural network trained with publicly accessible CT datasets. *Magn Reson Med.* (2022) 88:391–405. doi: 10.1002/mrm.29184
36. Hirtz S, Gonska T, Seydewitz HH, Thomas J, Greiner P, Kuehr J, et al. CFTR channel function in native human colon correlates with the genotype and phenotype in cystic fibrosis. *Gastroenterology.* (2004) 127:1085–95. doi: 10.1053/j.gastro.2004.07.006
37. Sommerburg O, Wielpütz MO, Trame J-P, Wuennemann F, Opdzaitze E, Stahl M, et al. Long-term effects of lumacaftor/ivacaftor on paranasal sinus abnormalities in infants and preschool children with cystic fibrosis. *Ann Am Thorac Soc.* (2020) 17:714–23. doi: 10.1513/AnnalsATS.201910-777OC
38. Wucherpfennig L, Wuennemann F, Eichinger M, Schmitt N, Seitz A, Baumann I, et al. Long-term effects of lumacaftor/ivacaftor on paranasal sinus abnormalities in children with cystic fibrosis detected with magnetic resonance imaging. *Front Pharmacol.* (2023) 14:1161891. doi: 10.3389/fphar.2023.1161891
40. Triphan SMF, Stahl M, Jobst BJ, Sommerburg O, Kauczor H-U, Schenk J-P, et al. Echo time-dependence of observed lung T1 in patients with cystic fibrosis and correlation with clinical metrics. *J Magn Reson Imaging.* (2020) 52:1645–54. doi: 10.1002/jmri.27271
41. Ciet P, Serra G, Bertolo S, Spronk S, Ros M, Fraioli F, et al. Assessment of CF lung disease using motion corrected PROPELLER MRI: a comparison with CT. *Eur Radiol.* (2016) 26:780–7. doi: 10.1007/s00330-015-3850-9
42. Dobritz M, Radkow T, Nittka M, Bautz W, Fellner FA. VIBE mit paralleler Akquisitionstechnik—eine neue Möglichkeit der dynamischen kontrastverstärkten MRT der Leber. *RöFo.* (2002) 174:738–41. doi: 10.1055/s-2002-32223
43. Semelka RC, Kelekis NL, Thomasson D, Brown MA, Laub GA. HASTE MR imaging: description of technique and preliminary results in the abdomen. *J Magn Reson Imaging.* (1996) 6:698–9. doi: 10.1002/jmri.1880060420
44. Stahl M, Wielpütz MO, Ricklefs I, Dopfer C, Barth S, Schlegtehdal A, et al. Preventive inhalation of hypertonic saline in infants with cystic fibrosis (PREIS). A randomized, double-blind, controlled study. *Am J Respir Crit Care Med.* (2019) 199:1238–48. doi: 10.1164/rccm.201807-1203OC
45. Robinson J. Likert scale In: AC Michalos, editor. *Encyclopedia of Quality of Life and Well-Being Research*. Dordrecht: Springer (2014). 3620–1.
46. Dice LR. Measures of the amount of ecologic association between species. *Ecology.* (1945) 26:297–302. doi: 10.2307/1932409
47. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. Sci Py 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods.* (2020) 17:261–72. doi: 10.1038/s41592-019-0686-2
48. Kirch W ed. Pearson’s correlation coefficient. In: *Encyclopedia of public health*. Dordrecht: Springer (2008).
49. Yang J, Zhang Z, Gong Y, Ma S, Guo X, Yang Y, et al. (2022). Do deep neural networks always perform better when eating more data? Available online at: <http://arxiv.org/pdf/2205.15187.pdf>
50. González Ballester MA, Zisserman AP, Brady M. Estimation of the partial volume effect in MRI. *Med Image Anal.* (2002) 6:389–405. doi: 10.1016/S1361-8415(02)00061-0
51. Maier-Hein L, Reinke A, Godau P, Tizabi MD, Buettner F, Christodoulou E, et al. Metrics reloaded: recommendations for image analysis validation. *Nat Methods.* (2024) 21:195–212. doi: 10.1038/s41592-023-02151-z
52. Weinheimer O, Konietzke P, Heussel CP, Kauczor H-U, Robinson TE, Galban CJ, et al. (2019). “Improving pulmonary lobe segmentation on expiratory CTs by using aligned inspiratory CTs” in *Medical Imaging 2019: Computer-Aided Diagnosis*, San Diego, California, United States. Bellingham, USA: SPIE. p. 126 (Progress in biomedical optics and imaging; vol. 20, no. 48). February 17–20, 2019.
53. Leewiwatwong S, Lu J, Dummer I, Yarnall K, Mummy D, Wang Z, et al. Combining neural networks and image synthesis to enable automatic thoracic cavity segmentation of hyperpolarized 129Xe MRI without proton scans. *Magn Reson Imaging.* (2023) 103:145–55. doi: 10.1016/j.mri.2023.07.001
54. Astley JR, Biancardi AM, Hughes PJC, Marshall H, Smith LJ, Collier GJ, et al. Large-scale investigation of deep learning approaches for ventilated lung segmentation using multi-nuclear hyperpolarized gas MRI. *Sci Rep.* (2022) 12:10566. doi: 10.1038/s41598-022-14672-2
55. Jin L, Ma Z, Li H, Gao F, Gao P, Yang N, et al. Interobserver agreement in automatic segmentation annotation of prostate magnetic resonance imaging. *Bioengineering.* (2023) 10. doi: 10.3390/bioengineering10121340 Available at: <https://www.mdpi.com/2306-5354/10/12/1340>
56. Liu Z, Tong L, Chen L, Jiang Z, Zhou F, Zhang Q, et al. Deep learning based brain tumor segmentation: a survey. *Complex Intell Syst.* (2023) 9:1001–26. doi: 10.1007/s40747-022-00815-5
57. Gandotra S, Kumar Y, Modi N, Choi J, Shafi J, Ijaz MF. Comprehensive analysis of artificial intelligence techniques for gynaecological cancer: symptoms identification, prognosis and prediction. *Artif Intell Rev.* (2024) 57. doi: 10.1007/s10462-024-10872-6
58. Tustison NJ, Qing K, Wang C, Altes TA, Mugler JP. Atlas-based estimation of lung and lobar anatomy in proton MRI. *Magn Reson Med.* (2016) 76:315–20. doi: 10.1002/mrm.25824