Journal of Cystic Fibrosis 24 (2025) 970–978

Contents lists available at ScienceDirect

Journal of Cystic Fibrosis

journal homepage: www.elsevier.com/locate/jcf

Original Article

Validation of an artificial intelligence-based automated PRAGMA and mucus plugging algorithm in pediatric cystic fibrosis☆

Pranali Raut [a,b,*], Yuxin Chen [a,b], Ahmad Taleb [a], Merlijn Bonte [a], Eleni Rosalina Andrinopoulou [c,d], Pierluigi Ciet [a,b,e], Jean-Paul Charbonnier [f], Claire E. Wainwright [g], Harm Tiddens [a,b,f], Daan Caudri [a,b]

[a] Department of Pediatric Pulmonology and Allergology, Erasmus MC -Sophia Children's Hospital, Rotterdam, The Netherlands
[b] Department of Radiology and Nuclear Medicine, Erasmus MC -Sophia Children's Hospital, Rotterdam, The Netherlands
[c] Department of Biostatistics, Erasmus MC, Rotterdam, The Netherlands
[d] Department of Epidemiology, Erasmus MC, Rotterdam, The Netherlands
[e] Department of Radiology, University of Cagliari, Italy
[f] Thirona B.V., Nijmegen, Netherlands
[g] Child Health Research Centre, University of Queensland, South Brisbane, Queensland, Australia

ARTICLE INFO

Keywords:
Artificial Intelligence
Cystic fibrosis
Computed tomography
PRAGMA
Quantification
Validation

ABSTRACT

Background: PRAGMA-CF is a clinically validated visual chest CT scoring method, quantifying relevant components of structural airway damage in CF. We aimed to validate a newly developed AI-based automated PRAGMA-AI and Mucus Plugging algorithm using the visual PRAGMA-CF as reference.

Material and Methods: The study included 363 retrospective chest CT's of 178 CF patients (100 New-Zealand and Australian, 78 Dutch) with at least one inspiratory CT matching the image selection criteria. Eligible CT scans were analyzed using visual PRAGMA-CF, automated PRAGMA-AI and Mucus Plugging algorithm. Outcomes were compared using descriptive statistics, correlation, intra- and interclass correlation and Bland-Altman plots. Sensitivity analyses evaluated the impact of disease severity, study cohort, number of slices and convolution kernel (soft vs. hard).

Results: The algorithm successfully analyzed 353 (97 %) CT scans. A strong correlation between the methods was found for %bronchiectasis (%BE) and %disease (%DIS), but weak for %Airway wall thickening (%AWT). The automated Mucus plugging outcomes showed strong correlation with visual %mucus plugging (%MP). ICC's between visual and automated sub-scores witnessed average agreement for %BE and %DIS, except for %AWT which was weak. Sensitivity analyses revealed that convolution kernel did not affect the correlation between visual and automated outcomes, but harder kernels yielded lower disease scores, especially for %BE and %AWT.

Conclusion: Our results show that AI-derived outcomes are not identical to visual PRAGMA-CF scores in size, but strongly correlated on measures of bronchiectasis, bronchial-disease and mucus plugging. They could therefore be a promising alternative for time-consuming visual scoring, especially in larger studies.

Abbreviations

A	Adjacent artery to the bronchi in the measurement
AI	Artificial Intelligence
AT	Atelectasis
AWT	Airway wall thickening
BA	Bronchial-Artery analysis
B_{in}	Bronchial inner diameter
B_{oa}	Bronchial outer area
B_{out}	Bronchial outer diameter
B_{wa}	Bronchial wall area
B_{wt}	Bronchial wall thickness
B_{out}/A	Bronchiectasis
B_{wt}/A	Bronchial wall thickening (calculated using artery)

☆ **Institution Information**, Erasmus MC- Sophia Children's Hospital, Dr. Molewaterplein 40, 3015 GD Rotterdam, The Netherlands

* Corresponding author at: Erasmus MC- Sophia Children's Hospital, Dr. Molewaterplein 40, 3015 GD Rotterdam, The Netherlands.
E-mail address: p.raut@erasmusmc.nl (P. Raut).

https://doi.org/10.1016/j.jcf.2025.08.003
Received 18 March 2025; Received in revised form 6 August 2025; Accepted 7 August 2025
Available online 21 August 2025

B_{wa}/B_{oa}	Bronchial wall thickening (calculated using bronchus itself)
BE	Bronchiectasis
BMI	Body mass index
CF	Cystic Fibrosis
COPD	Chronic Obstructive Pulmonary Disease
CT	Computed Tomography
CF-CT	Cystic Fibrosis–Specific Computed Tomography Scoring
CFTR	Cystic fibrosis transmembrane conductance regulator
DIS	Disease
ECFS	European cystic fibrosis society
ICC	Intra- and Inter-class correlation coefficient
ILD	Interstitial lung disease
MP	Mucus plugging
PCD-CT	photon-counting detector CT
PRAGMA-CF	Perth-Rotterdam Annotated Grid Morphometric Analysis for cystic fibrosis
TIB	Tree-in-bud

1. Introduction

Cystic Fibrosis (CF) is a genetic, multi-systemic disorder primarily affecting the respiratory system, causing recurrent infection, chronic inflammation, and progressive structural lung damage starting from early childhood [1]. Standard reference tests, such as pulmonary function test (PFT) are often insensitive to early structural lung damage and fail to localize the underlying abnormalities [2]. Conversely, medical imaging, an essential tool used in routine follow-up of CF patients, can accurately visualize the extent and location of minimal structural lung disease [3]. Computed tomography (CT) is currently considered as the gold standard to assess structural lung disease, due to its high spatial resolution and rapid acquisition protocol [4]. Recent advancements in CT scanner technology have enabled ultra-low dose radiation protocols, allowing safe monitoring of CF lung disease progression [5].

Several semi-quantitative CT-based scoring methods, such as Bhalla score [6], Brody score [7], and Cystic Fibrosis–Specific CT Score (CF-CT) [7,8] have been initially developed to objectively quantify the extent of abnormalities like bronchiectasis, airway wall thickening, mucus plugging, atelectasis, and trapped air. These scores were shown to detect clinically relevant structural abnormalities, but were relatively insensitive for longitudinal monitoring of slowly progressing lung disease. The Perth-Rotterdam Annotated Grid Morphometric Analysis in CF (PRAGMA-CF) was developed for this purpose, and has shown higher sensitivity in identifying minor progression and monitoring structural lung disease [9]. The PRAGMA-CF scoring is validated for its repeatability [10], correlation with other clinical markers of disease [11] and serving as an outcome measure in clinical trials [12]. However, its application in clinical practice is limited due to its observer-dependence, time--consuming nature, and associated cost. These challenges could be addressed by utilizing Artificial Intelligence (AI)-based algorithms for the automated quantification of abnormalities. Several CF specific AI-algorithms were recently developed, including NOVAA-CT [13] and the LungQ platform, the latter of which includes a Bronchial-Artery (BA) algorithm, a Mucus Plugging algorithm, and VERA analysis for air-trapping [14]. The BA algorithm was adapted to also produce outcomes similar to visual PRAGMA-CF scores (PRAGMA-AI). Mucus plugging is not included in that PRAGMA-AI algorithm, but the separate Mucus Plugging algorithm is able to quantify the number and volume of mucus plugs obstructing the detectable airways.

This study aims to evaluate the efficiency of AI-based algorithms by comparing their outcomes with the clinically validated visual PRAGMA-CF outcomes and analyze the factors that could affect the automated quantification and therefore its association with the visual method. Our goal is to explore the feasibility of using AI-based scoring in clinical research and care settings. While we anticipated discrepancies in the outcomes due to differences in design and outcome units of the methods, we did expect a correlation in identifying key structural lung disease characteristics. Furthermore, automated algorithms could provide observer-independent reproducible outcomes, with the capability for large-scale data processing, positioning them as promising tools for broader clinical application [15].

2. Material and methods

2.1. Study population and Image acquisition

The retrospective study included two pediatric cohorts of patients with a confirmed diagnosis of CF and at least one routine CT scan available. The first cohort of New Zealand and Australian CF-FAB multicenter study included 100 children aged between 9—18 years, with total 186 inspiratory CT scans (FAB) [16]. The second cohort of Dutch Erasmus MC Sophia, Rotterdam institute, included 78 children aged between 6—18 years, with total 177 inspiratory routine CT scans (SOPHIA). FAB cohort chest CT's were acquired with technician-controlled breathing technique, while for SOPHIA cohort spirometry-controlled breathing standards were used. Patient demographics and image characteristics of both cohorts are summarized in Table 1 and **Supplementary Material**. None of the patients included in this study were on Elexacaftor/tezacaftor/ivacaftor (ETI) modulator therapy.

2.2. Image analysis

The technical requirements of automated analysis were, an axial inspiratory CT scans with a slice thickness $\leq$ 1.5 mm, free of large artifacts or missing slices. Both soft and hard convolution kernel reconstructions were accepted, and their effects were evaluated. All scans were assessed using visual PRAGMA-CF, automated PRAGMA-AI and Mucus Plugging scoring algorithm.

2.3. Visual analysis

The visual PRAGMA-CF score is a grid-based morphometric score, where 10 randomly selected equally spaced slices are overlaid with a squared grid. Abnormalities within each grid cell are scored hierarchically for bronchiectasis (BE), mucus plugging (MP), airway wall thickening (AWT), atelectasis (AT), and normal lung and the results are computed as a percentage of total volume (**Supplementary Material**). Atelectasis was rare in this cohort and was not included in the %DIS

Table 1
Patient and imaging related characteristic of the study population.

Variable	Specifications	
Demographics		
Cohort	FAB	SOPHIA
Country	Australia New Zealand	The Netherlands
Patients	100	78
Median Age (range)	13.7 (9-18) years	11.0 (6-18) years
Functional Parameters		
Median %FEV_1	86.27 (73.38 – 94.60)	102.12 (91.96 – 109.89)
Median %FVC	93.26 (82.72 – 104.18)	103.50 (97.06 – 111.42)
Median BMI	19.36 (17.67 – 20.88)	17.08 (15.70 – 19.04)
Imaging		
Imaging technique	Computed tomography	Computed tomography
Scanner vendor	SIEMENS, PHILIPS, GE, TOSHIBA	SIEMENS
Image type	Technician controlled inspiratory CT	Spirometry controlled inspiratory CT
Number of CT's available		
Total (any kernel)	186	177
hard kernel	34	3
soft kernel	63	0
Both hard and soft kernel	89	174

score. The FAB and SOPHIA datasets were scored by two trained, certified and experienced observers (Y. Chen and M. Bonte respectively) using the in-house developed SALDSegVol software, resulting in the sub-scores of %BE, %AWT, %MP, %AT, and the overall composite score % Disease (%DIS= %BE+ %MP+ %AWT). To evaluate reliability, intra-class correlation coefficients (ICC) were calculated by each observer through rescoring a random subset of 25 scans again at least one month after their first scoring. The inter-class correlation coefficient was computed using a randomly selected batch of 25 scans from the other dataset. The hardest kernel was chosen amongst multiple reconstructions if available, for optimal visual scoring of lung abnormalities.

2.4. Automated analysis

All scans were analyzed using the PRAGMA-AI algorithm of LungQ^{TM} platform, identifying %BE, %AWT, and %DIS. The PRAGMA-AI algorithm uses the base algorithm of previously validated LungQ-BA algorithm [14], a brief description of which can be found in **Supplementary material**. The measurements by the BA algorithm were validated against actual manual measurements performed on over 4500 individual airways in CF patients and healthy controls [14]. The algorithm computes the PRAGMA-AI outcomes, by overlaying each CT slice with a virtual square grid similar to visual scoring. Each grid cell is scored hierarchically making use of the BA algorithm and outcomes. Predefined thresholds are used to define thresholds for specific lung abnormalities, namely bronchiectasis (bronchus-artery ratio $>$ 1.3968), bronchial wall thickening (bronchial wall-artery ratio $>$ 0.1879), or other lung tissue/airways. Unlike manual method, the automated PRAGMA-AI provides the option to generate outcomes using 10 randomly selected slices, or by using all available slices. Mucus plugging and atelectasis assessment is not part of PRAGMA-AI algorithm. Instead, a separate Mucus Plugging algorithm was developed that scans the entire segmented airway tree and identifies total number and computes the corresponding volume (ml) of detectable mucus plugs causing airway obstructions. For the main analyses we made use of the results from all available slices and if multiple kernels were available for analysis, the softest convolution kernel appropriate for lung assessment was chosen as is recommended for the LungQ software.

2.5. Statistical analysis

The two-way mixed effect ICC's for intra- and inter-observer agreement below 0.5 was rated as poor, 0.5—0.75 as moderate, 0.75—0.9 as good, and above 0.9 as excellent [17]. Visual and automated outcomes were evaluated using descriptive statistics. The normality of the data was assessed using Shapiro-Wilk test and QQ plots. Visual and automated outcomes were then compared using Spearman correlation coefficient, where correlation lower than 0.2 was rated as very weak, 0.2—0.4 as weak, 0.4—0.6 as moderate, 0.6—0.8 as strong, and 0.8—1.0 as excellent [18]. Despite the outcomes not being directly comparable, we assessed their agreement using ICC's and Bland-Altman plots for outcomes with identical units. Sensitivity analyses were performed to assess the impact of multiple factors, such as the number of CT slices used (10 vs. all slices), type of kernel (hard vs. soft), cohort, and disease severity on the correlation and agreement statistics. Statistical analyses were conducted using R software (Version- 4.4.1).

3. Results

The ICC's, within and between two visual scorers for each sub-score are summarized in Table 2. Except for lower agreement in %AWT, all other sub-scores showed good to excellent agreement.

363 chest CT scans (FAB=186 [soft kernel: 82 %, hard kernel: 18 %] and SOPHIA=177 [soft kernel: 98 %, hard kernel: 2 %]) were scored visually and using automated algorithm of LungQ platform. The algorithm successfully analyzed 353 chest CT scans (97 %, FAB=181(97 %) and SOPHIA=172(97 %)). On average a standard CT was processed in about 30 minutes, while a photon counting CT scan took about 50 minutes to process using LungQ. Scans that failed during automated post-processing were excluded from the analysis. The reasons of failed scans are listed in the **Supplementary material**.

Table 2
Summary of Intra- and Inter-observer ICC with its corresponding 95 % confidence interval ($CI_{95\ \%}$).

Variable	Intra-class correlation coefficient scorer 1* [$CI_{95\ \%}$]	Intra-class correlation coefficient scorer 2* [$CI_{95\ \%}$]	Inter-class correlation coefficient between 2 scorer 1* and 2* [$CI_{95\ \%}$]
%BE	0.96 [0.92, 0.98]	0.98 [0.95, 0.99]	0.81 [0.21, 0.94]
%AWT	0.54 [0.21, 0.77]	0.63 [0.21, 0.85]	0.24 [-0.09, 0.55]
%MP	0.92 [0.83, 0.96]	0.96 [0.89, 0.99]	0.96 [0.91, 0.98]
%AT	0.84 [0.68, 0.93]	0.91 [0.72, 0.97]	0.77 [0.47, 0.90]
%DIS	0.99 [0.98, 1.00]	0.97 [0.93, 0.99]	0.95 [0.80, 0.98]

ICC for visual PRAGMA-CF scoring. Within and between the observers calculated based on random subset of 25 scans scored in Duplo. *: Scorer Y.C. in FAB cohort; #: Scorer M.B. in SOPHIA cohort; $: Scorers Y.C. & M.B compared to each other

Descriptive statistics of visual PRAGMA-CF, automated PRAGMA-AI and Mucus Plugging outcomes, computed for "combined dataset" (FAB+SOPHIA) as well as each cohort is summarized in Table 3. The combined dataset consisted of 353 CT scans in total, excluding unprocessed scans and those with missing visual PRAGMA-CF scores.

3.1. Comparison of visual vs automated CT outcomes in Combined dataset

The tests conducted to check the assumption of residuals revealed non-normal distributions for both visual and automated sub-scores of the combined dataset. Therefore, spearman correlation was used to determine the relationship between visual and automated outcomes. A strong positive correlation was observed between visual and automated scoring methods for %BE and %DIS (0.84; CI95 %: 0.81-0.87, $p<0.001$ and 0.82; CI95 %: 0.78-0.85, $p<0.001$, respectively). A weak correlation was observed for %AWT of 0.36 (CI95 %: 0.27-0.45, $p<0.001$). The comparison of visual %MP and automated number of mucus plugs as well as mucus volume showed strong correlations of 0.78 (CI95 %: 0.74-0.82, $p<0.001$) and 0.79 (CI95 %: 0.75-0.83, $p<0.001$), respectively. Scatterplots of these sub-scores are shown in Fig. 1.

ICC analysis and Bland-Altman plots were performed for the sub-scores with identical units in visual and automated analyses, namely %BE, %AWT, and %DIS. The ICC for %BE and %DIS indicated moderate agreement (0.69; $CI_{95\ \%}$: 0.55-0.78, and 0.59; $CI_{95\ \%}$: 0.43-0.70, respectively). A weak agreement was found for %AWT of 0.25 ($CI_{95\ \%}$: 0.15-0.35). Bland-Altman plots revealed that the limits of agreement were relatively wide for all outcomes (Fig. 2), and for %BE and %DIS, the difference between visual and automated outcomes increases when the average value increases. This is further supported by the correlation analysis and regression line plotted in Fig. 2A and C, revealing a pattern that automated scores tend to underestimate results compared to visual scores.

3.2. Sensitivity analyses

No meaningful differences were found when comparing automated PRAGMA-AI sub-scores, computed using 10 randomly selected slices (analogous to procedure in visual PRAGMA-CF) versus all available slices of CT scan (see **Supplementary material**). Therefore, automated results computed using all available slices were analysed and reported. To analyze the influence of convolution kernel, we compared automated outcomes generated using soft kernel CT scans with those generated

Table 3
Descriptives of visual and automated outcomes for FAB, SOPHIA and combined dataset.

Outcomes	Combined Dataset		FAB		SOPHIA	
	MEDIAN	IQR*	MEDIAN	IQR*	MEDIAN	IQR*
Visual variables						
%BE	3.63	1.68 – 7.12	4.90	2.25 – 9.90	2.89	1.39 – 4.65
%AWT	0.62	0.20 – 1.35	0.46	0.14 – 0.73	0.99	0.35 – 2.23
%DIS	5.30	3.03 – 9.39	6.01	3.05 – 12.44	4.82	3.01 – 7.12
%MP	0.00	0.00 – 0.87	0.19	0.00 – 1.30	0.00	0.00 – 0.57
%AT	0.13	0.00 – 0.51	0.06	0.00 – 0.49	0.17	0.00 – 0.55
Automated variables						
%BE	3.00	1.35 – 5.24	3.47	1.37 – 6.00	2.57	1.25 – 4.58
%AWT	1.02	0.45 – 2.02	1.36	0.60 – 2.34	0.69	0.31 – 1.74
%DIS	4.32	2.42 – 6.76	5.24	2.72 – 8.03	3.76	2.23 – 5.88
Mucus plugs	1.00	0.00 – 10.00	3.00	0.00 – 17.00	0.00	0.00 – 4.00
Mucus Volume	0.03	0.00 – 0.59	0.12	0.00 – 1.17	0.00	0.00 – 0.21

IQR*: Interquartile range

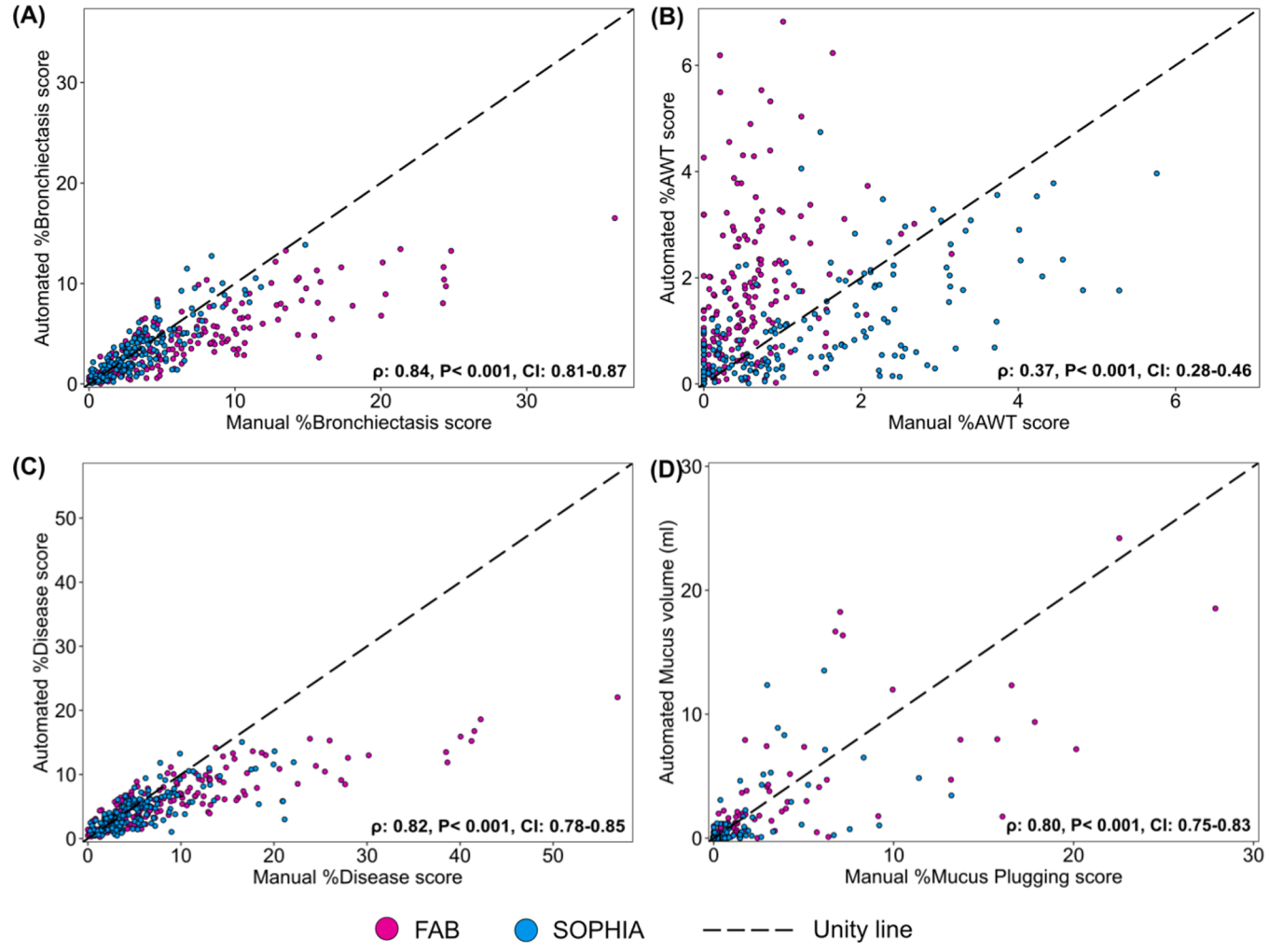

Fig. 1. The scatterplots of visual vs automated measurements for (A) Bronchiectasis, (B) Airway wall thickening, (C) Disease and (D) %Mucus plugging vs Mucus Volume (ml).
353 CT scans of Combined dataset (FAB and SOPHIA) were used for the analysis where colors of plot indicates the clinical cohorts used, where ρ: spearman correlation coefficient, p-value indicates null hypothesis of correlation is zero, $CI_{95\ \%}$: 95 % confidence interval

from hard kernel CT scans. Out of each 174 soft and hard kernel CT scans of SOPHIA, 162 successfully post-processed scans common between both kernels were selected for this sub-study. Spearman correlation analysis showed good to excellent correlation for all sub-scores (Table 4 and Fig. 3A, B). However, ICC analysis revealed only a moderate agreement for %BE and %AWT, while good agreement for %DIS and excellent agreement for mucus plugs and mucus volume were noticed (Table 4).

Bland-Altman plots showed similar pattern to the previous comparison of visual and automated outcomes. Again, high correlation was observed between the sub-scores, but the actual values were not identical. Specifically, using hard kernel CT systematically showed lower

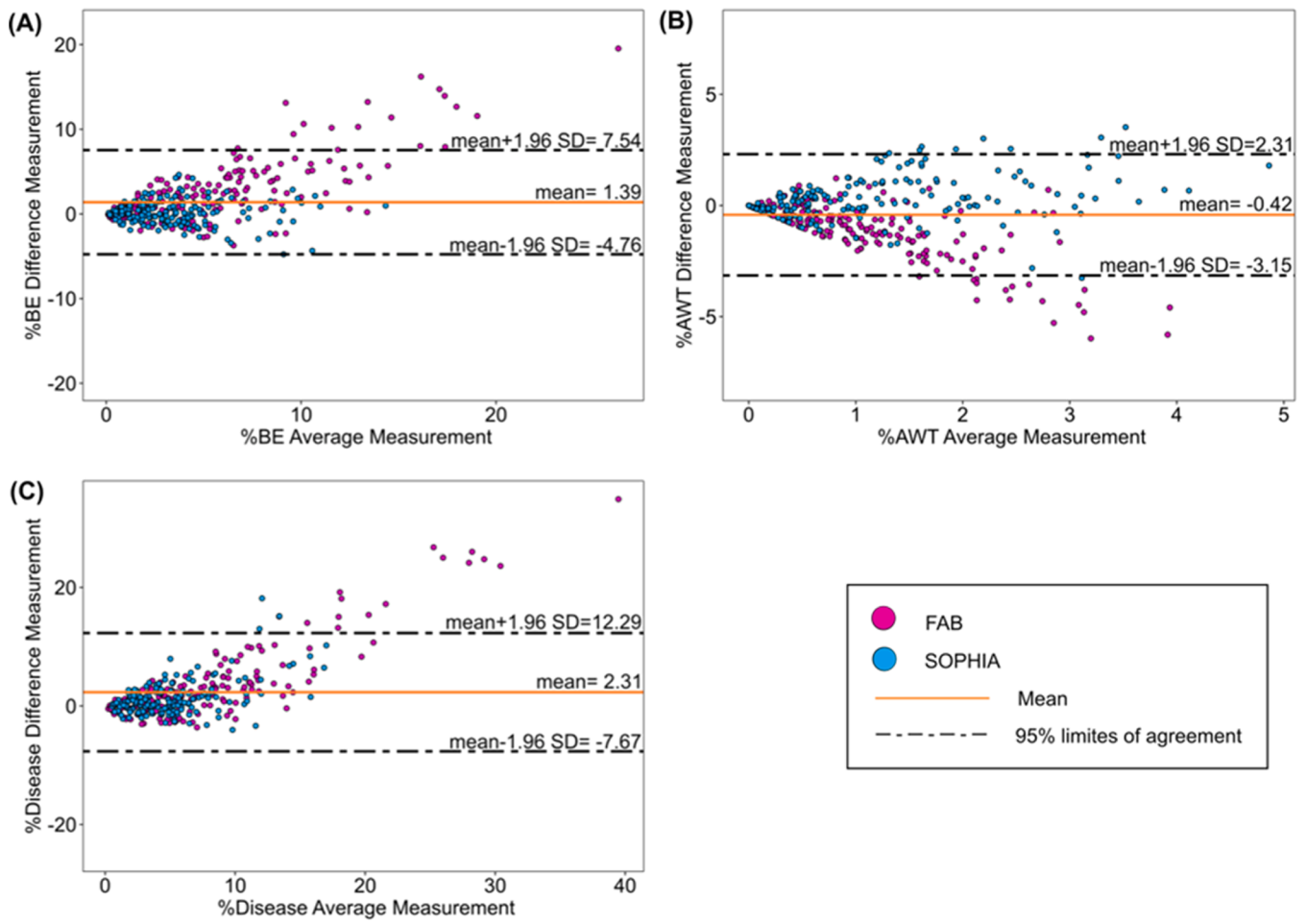

Fig. 2. The Bland-Altman plots drawn for sub-scores of (A) %Bronchiectasis, (B) % Airway wall thickening, and (C) %Disease.
The difference between visual and automated outcomes are plotted against mean of visual and automated outcomes for each sub-score for 353 CT scans of Combined dataset, where colors indicate clinical cohorts used.

Table 4
Correlation and ICC analysis comparing automated outcomes of hard kernel vs soft kernels.

Variable	Spearman correlation			Inter-class correlation coefficient	
	ρ	$CI_{95\ \%}$	p-value	ICC	$CI_{95\ \%}$
%BE	0.87	0.83-0.91	<0.001	0.51	-0.07-0.78
%AWT	0.81	0.75-0.86	<0.001	0.55	0.01-0.76
%DIS	0.91	0.87-0.93	<0.001	0.83	0.23-0.93
Mucus plugs	0.96	0.94-0.97	<0.001	0.99	0.98-0.99
Mucus volume (ml)	0.96	0.95-0.97	<0.001	0.99	0.98-0.99

Summary of correlation analysis and Inter-class coefficient (ICC) analysis computed for 162 CT scans of SOPHIA comparing automated outcomes generated using soft kernel CT scans and hard kernel CT scans, where ρ: correlation coefficient, ICC: Inter-class correlation coefficient and $CI_{95\ \%}$: 95 % confidence interval

automated outcomes compared to soft kernel CT (Fig. 3). The hard kernel overestimated %AWT, while %BE was underestimated, in comparison to manual scoring. However, a smaller impact by kernels was noticed on the combined outcome %DIS. A detailed representation of these findings comparing, visual PRAGMA-CF outcomes with automated outcomes computed using hard kernel, and visual PRAGMA-CF outcomes with automated outcomes computed using soft kernel, is provided in the **Supplementary material**.

Additional analyses looking at the effects of disease severity and study cohort revealed no meaningful effects on the relations between visual and automated results (data not shown).

4. Discussion

In this study, we compared the results of a newly developed automated PRAGMA-AI and Mucus Plugging algorithms with the clinically validated visual PRAGMA-CF outcomes. The automated algorithm successfully processed 97 % of applied CT scans, showing its feasibility in a real-world clinical setting. We found strong to excellent correlations between visual and automated sub-score for %BE and %DIS, while the ICC and Bland-Altman analysis indicated only a moderate agreement. Interestingly, the Mucus Plugging algorithm showed excellent correlation with visual %MP, despite having different outcomes units, suggesting it can effectively capture clinically relevant mucus plugs.

A key finding of our study is the strong correlation we found between the methods for assessment of %BE and %DIS. Bronchiectasis is the most clinically relevant marker of progressive CF lung disease, associated with exacerbations, reduced lung function, and survival [11,16,19]. Our findings indicate that PRAGMA-AI can indeed detect the signal of this key outcome similar to visual PRAGMA-CF score. However, the moderate agreement between the methods highlights that they should not be used interchangeably.

Contrary to %BE and %DIS, a weak correlation was found for %AWT between the methods. The visual assessment of AWT is challenging, subjective and prone to considerable variability, both within and between observers [12,20]. This is also clearly reflected in the low intra-

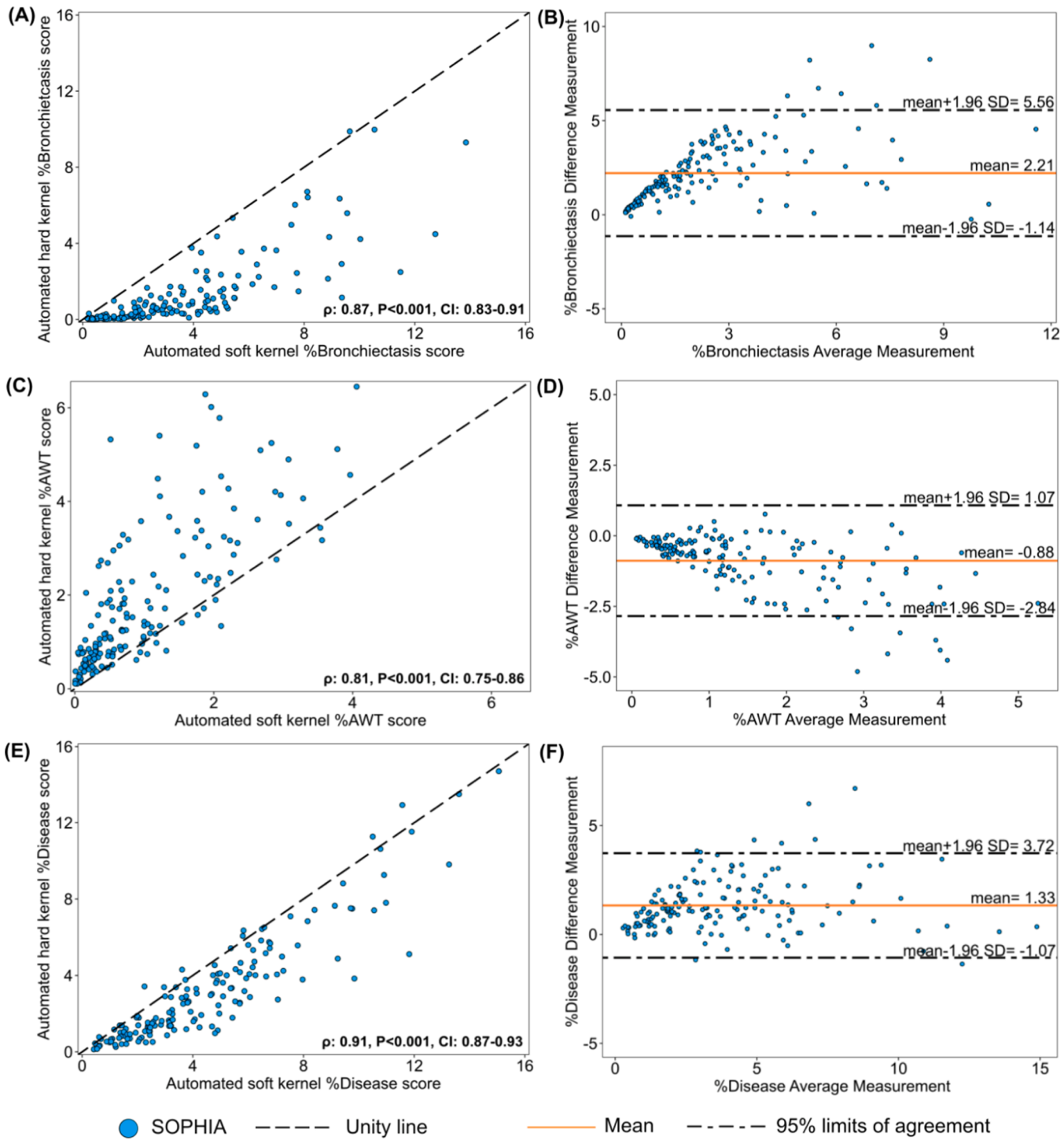

Fig. 3. Scatter- and Bland-Altman plots comparing hard vs. soft kernel automated outcomes for %Bronchiectasis, %Airway wall thickening and %Disease. 162 CT scans of SOPHIA cohort post-processed successfully, available with both hard and soft kernels were chosen for the sub-study, where ρ: spearman correlation coefficient, p-value indicates null hypothesis of correlation is zero, and $CI_{95\ \%}$: 95 % confidence interval

and interobserver ICC's given in Table 2. In this context the visual %AWT assessment may not at all be considered the ideal reference test to compare the automated tool with. The automated %AWT on the other hand, by design, offers a perfect repeatability. But this does imply it is more accurate in detecting true AWT and estimates can be affected by factors like kernel type and breathing standardization. However, both scoring methods are hierarchical in nature and prioritize BE over AWT as clinically BE is considered more permanent and relevant structural lung damage than AWT, which leads to subsequent underestimation of AWT. We found excellent correlation between visual and automated outcomes for %DIS, despite %AWT sub-score is incorporated in it. This suggests that AWT only contributes a small portion to the total %DIS. Similar observations were made in other studies where CF is primarily characterized by BE rather than AWT [21,22].

The Mucus Plugging algorithm utilizes airway segmentation to detect localized mucus plugs causing complete occlusions in the lumen of detected airway generations [23]. This approach distinctly differs from the visual %MP sub-score, which primarily captures peripheral

tree-in-bud patterns and some larger central mucus plugs. Despite these differences, we found excellent correlations between the two assessment methods of mucus plugging. We hypothesize that the majority of peripheral tree-in-bud patterns often associates with centrally localized mucus plugs obstructing the mucus and air flow in all higher airway generations that branch out from that location. While visually identifying these plugs may be possible, it would be extremely time-consuming. Recently, there has been increasing focus on the role of mucus in the pathogenesis of bronchiectasis, both in CF and non-CF [24,25]. Mucus may not just be a marker of inflammation but could also be a critical, causal component in the vicious vortex leading to bronchiectatic airway damage [26]. Consequently, MP may serve as an essential sub-score detecting active infection and inflammation [27], particularly in patients with clinically symptomatic bronchiectatic airway disease [27,28]. Unlike BE, MP likely reflects a stages of disease where implemented interventions could still prevent progression to permanent structural lung damage. Our findings suggest that the automated Mucus plugging algorithm is as effective as %MP of visual PRAGMA-CF, in detecting clinically relevant occlusions in the airways, even though the outcome units of the two methods are fundamentally different.

As part of a sensitivity analysis, we compared automated sub-scores generated using 10 randomly selected slices and all available slices with visual sub-scores. We found no clear difference, supporting the assumption that 10 randomly selected slices provide a reliable representation of overall structural lung damage [9]. However, theoretically, using all available slices minimizes the risk of missing small, localized but clinically relevant abnormalities, making this approach preferable for more comprehensive analysis. We also observed differences in automated outcomes depending on whether hard or soft CT reconstruction kernels were used. Several studies have documented the impact of kernel selection on lung abnormalities assessment using automated algorithms [29,30]. The sensitivity analysis comparing automated outcomes of soft and hard kernels, showed good to excellent correlation, but revealed substantial variation in the prevalence of detected abnormalities, particularly in %BE and %AWT. Importantly, using soft kernels resulted in higher estimates of %BE compared to hard kernels (Figure 3A), but lower estimates of %AWT compared to hard kernels (Figure 3C). The comparison of soft kernel sub-scores and hard kernel sub-scores with its counterpart manual sub-scores showed similar trend. The soft kernel sub-scores showed proportional estimates to manual sub-scores, while hard kernels underestimated the %BE sub-score and overestimated the %AWT sub-score (see **Supplementary material**). Our findings support the notion that the used convolution kernel can play an important role and should be chosen considerately to avoid introducing bias. It should however also be noted that similar biases may occur in manual analysis, albeit much more laborious to investigate. In clinical practice, optimal reconstruction kernels may not always be available; our findings highlight the importance to at least be consistent in kernel selection, particularly when tracking disease progression or treatment response within patients. This also stresses the importance of standardization of CT image acquisition protocols in research and clinical care.

Our study has several notable strengths. It includes multi-center data that encompasses a diverse range of disease severities and incorporates CT scans acquired using various vendors and protocols, enhancing generalizability of our findings. The large sample size enabled us to examine the effects of various factors affecting automated scoring mechanism, most notably reconstruction kernel. Highly experienced scorers performed the visual scoring, limiting the variability and ensuring an optimal reference for comparison with the automated outcomes. However, there are several limitations to our study. Although visual PRAGMA-CF is used as the gold standard, may not be a perfect reference method for the comparison, and the observed differences between visual and automated scores could in fact be due to misclassifications by either method. Additionally, the automated algorithm did not incorporate all sub-scores, (e.g. atelectasis), and used a different method for assessing mucus plugging, making direct comparisons challenging. The algorithm is also limited by its inability to quantify abnormalities in the most peripheral airways. We found that the automated algorithm is prone to the effects reconstruction kernel used in CT acquisition and the fact that soft kernels were not available for all CT scans likely influenced the outcomes. Finally, it should be noted that this analysis only investigates the PRAGMA-AI tool in relation to the well validated manual PRAGMA-CF tool. Other AI-supported tools to quantify CF structural lung disease have been developed and may have benefits over PRAGMA-AI such as BA algorithm, NOVAA-CT etc. [13,14]. Direct comparison of these outcomes with PRAGMA-CF is less relevant due to fundamental differences in the algorithm's underlying principles, the types of abnormalities they detect, and the units in which outcomes are reported.

What are the clinical implications of our findings? Our analysis suggests that the PRAGMA-AI and Mucus Plugging algorithms are capable of effectively quantifying the most relevant abnormalities of CF, similar to visual PRAGMA-CF. In large clinical trials, registry studies, and in busy routine clinical care where visual scoring seems impractical, the automated algorithm could serve as a quick alternative, alleviating the burden of manual labor, while providing consistent, observer-independent quantified outcomes. Especially in the post-CFTR modulator era, the need for a sensitive clinical measures to detect and monitor the earliest signs of CF lung disease progression has become paramount, as traditional tests like flow-volume curves and sputum cultures have become less sensitive and more challenging [31]. Over the last decade, chest CT has become a common modality for routine monitoring and assessment of the progression of structural lung abnormalities in many CF centers in Europe and Australia [4,20]. Using automated algorithms like PRAGMA-AI and Mucus plugging can pave the way to leverage the information captured in these routine CT scans to advance our understanding of structural lung damage in CF. We aim to use these algorithms in the ENRICH study, adding automated quantified structural lung outcomes of 5.000-10.000 chest CT scans to the ECFS patient registry. The results from these analyses will add to our knowledge and understanding of the ability of PRAGMA-AI and other LungQ outcomes to detect structural lung abnormalities in PwCF, to monitor disease progression over time and to assess treatment effects of ETI in this population.

In conclusion, the newly developed PRAGMA-AI and Mucus Plugging algorithms can effectively quantify structural lung abnormalities in children with CF and correlate strong with outcomes of the clinically validated visual PRAGMA-CF score. The excellent correlation implies that the automated algorithm can successfully replicate the key outcomes of visual method, while the moderate agreement indicates that they should not be considered equivalent or used interchangeably. For the future large clinical trials and registry studies, automated quantification algorithms are likely to play a pivotal role in assessing CF structural lung disease.

Conflict of interest

The co-author H. Tiddens is emeritus professor at Erasmus Medical center and chief medical officer of Thirona, the company providing LungQ^{TM} software. He is also shareholder in Thirona and has received grants from Insmed and Boehringer Ingelheim to conduct the research. He is sponsored by Vertex Pharmaceuticals to delivered Faculty ADVANCE course. The co-author J.P. Charbonnier is a chief innovation officer at Thirona. The co-author C.E. Wainwright has participated in clinical trials sponsored by Vertex Pharmaceuticals, educational meetings and advisory boards for Vertex Pharmaceuticals and her institution has received funding for her participation in these activities and she has received travel and accommodation to investigator meetings. CF FAB was funded by NHMRC grant 1044829. The remaining co-authors have no declaration of interest to disclose.

Funding

This study was covered under the 2021 Out of Cycle CFF grant entitled "ENRICH the ECFS Patient Registry with Structural Lung data from Chest CT scans", 1-10-2022 – 30-9-2026. (Award number: 003138121).

Ethics statement

This FAB study involves human participants from multiple centers and was approved by the Medical Ethical Review Committee of respective participating institutes. The informed consent forms were signed by the participants before taking part in the study. The data from the Sophia Children's Hospital was collected under a waiver for METC approval and all participants signed informed consent for the use of their pseudonymized data.

Data availability statement

Only computed outcomes of the visual and automated assessment can be made available on reasonable request. CT scans of the of the patients will not be shared.

CRediT authorship contribution statement

Pranali Raut: Data curation, Formal analysis, Investigation, Methodology, Project administration, Validation, Visualization, Writing – original draft, Writing – review & editing. **Yuxin Chen:** Methodology, Writing – review & editing. **Ahmad Taleb:** Methodology, Writing – review & editing. **Merlijn Bonte:** Methodology, Writing – review & editing. **Eleni Rosalina Andrinopoulou:** Formal analysis, Writing – review & editing. **Pierluigi Ciet:** Resources, Writing – review & editing. **Jean-Paul Charbonnier:** Software, Writing – review & editing. **Claire E. Wainwright:** Resources, Writing – review & editing. **Harm Tiddens:** Conceptualization, Writing – review & editing. **Daan Caudri:** Conceptualization, Funding acquisition, Investigation, Project administration, Resources, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Co-author H. Tiddens is emeritus professor at Erasmus Medical center. He is also chief medical officer and stakeholder of Thirona, the company providing LungQ^{TM} software. The co-author J.P. Charbonnier is a chief innovation officer at Thirona.

Acknowledgment

We would like to acknowledge the FAB study research group for sharing their data and granting permission to use CT images in our study and the Erasmus MC Graduate School for providing statistical training that supported the methodological aspects of this work.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jcf.2025.08.003.

References

[1] Rosenfeld M, Gibson RL, McNamara S, et al. Early pulmonary infection, inflammation, and clinical outcomes in infants with cystic fibrosis. Pediatr Pulmonol 2001;32:356–66. https://doi.org/10.1002/ppul.1144.
[2] de Jong PA, Nakano Y, Lequin MH, et al. Progressive damage on high resolution computed tomography despite stable lung function in cystic fibrosis. Eur Respir J 2004;23:93–7. https://doi.org/10.1183/09031936.03.00006603.
[3] Murphy KP, Maher MM, O'Connor OJ. Imaging of cystic fibrosis and pediatric bronchiectasis. Am J Roentgenol 2016;206:448–54. https://doi.org/10.2214/AJR.15.14437.
[4] Rybacka A, Karmelita-Katulska K. The role of computed tomography in monitoring patients with cystic fibrosis. Pol J Radiol 2016;81:141–5. https://doi.org/10.12659/PJR.896051.
[5] O'Connor OJ, Vandeleur M, McGarrigle AM, et al. Development of low-dose protocols for thin-section CT assessment of cystic fibrosis in pediatric patients. Radiology 2010;257:820–9. https://doi.org/10.1148/radiol.10100278.
[6] Bhalla M, Turcios N, Aponte V, et al. Cystic fibrosis: scoring system with thin-section CT. Radiology 1991;179:783–8. https://doi.org/10.1148/radiology.179.3.2027992.
[7] Brody AS, Kosorok MR, Li Z, et al. Reproducibility of a scoring system for computed tomography scanning in cystic fibrosis. J Thorac Imaging 2006;21:14. https://doi.org/10.1097/01.rti.0000203937.82276.ce.
[8] Brody AS, Klein JS, Molina PL, et al. High-resolution computed tomography in young patients with cystic fibrosis: distribution of abnormalities and correlation with pulmonary function tests. J Pediatr 2004;145:32–8. https://doi.org/10.1016/j.jpeds.2004.02.038.
[9] Rosenow T, Oudraad MCJ, Murray CP, et al. PRAGMA-CF. A quantitative structural lung disease computed tomography outcome in young children with cystic fibrosis. Am J Respir Crit Care Med 2015;191:1158–65. https://doi.org/10.1164/rccm.201501-0061OC.
[10] Kuo W, Andrinopoulou E-R, Perez-Rovira A, et al. Objective airway artery dimensions compared to CT scoring methods assessing structural cystic fibrosis lung disease. J Cystic Fibrosis 2017;16:116–23. https://doi.org/10.1016/j.jcf.2016.05.015.
[11] Turkovic L, Caudri D, Rosenow T, et al. Structural determinants of long-term functional outcomes in young children with cystic fibrosis. Eur Respir J 2020;55. https://doi.org/10.1183/13993003.00748-2019.
[12] Tiddens HAWM, Chen Y, Andrinopoulou E-R, et al. The effect of inhaled hypertonic saline on lung structure in children aged 3–6 years with cystic fibrosis (SHIP-CT): a multicentre, randomised, double-blind, controlled trial. Lancet Respir Med 2022;10:669–78. https://doi.org/10.1016/S2213-2600(21)00546-4.
[13] Hadj Bouzid AI, Bui S, Benlala I, et al. Artificial intelligence-driven volumetric CT outcome score in cystic fibrosis: longitudinal and multicenter validation with/without modulators treatment. Eur Radiol 2025;35:815–27. https://doi.org/10.1007/s00330-024-11019-5.
[14] Lv Q, Gallardo-Estrella L, Andrinopoulou E-R, et al. Automatic analysis of bronchus-artery dimensions to diagnose and monitor airways disease in cystic fibrosis. Thorax 2024;79:13–22. https://doi.org/10.1136/thorax-2023-220021.
[15] Handa T. The potential role of artificial intelligence in the clinical practice of interstitial lung disease. Respir Investig 2023;61:702–10. https://doi.org/10.1016/j.resinv.2023.08.006.
[16] Begum N, Byrnes CA, Cheney J, et al. Factors in childhood associated with lung function decline to adolescence in cystic fibrosis. J Cystic Fibrosis 2022;21:977–83. https://doi.org/10.1016/j.jcf.2022.03.008.
[17] Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016;15:155–63. https://doi.org/10.1016/j.jcm.2016.02.012.
[18] Chen Y, Lv Q, Andrinopoulou E-R, et al. Automatic bronchus and artery analysis on chest computed tomography to evaluate the effect of inhaled hypertonic saline in children aged 3-6 years with cystic fibrosis in a randomized clinical trial. J Cyst Fibrosis 2023;22:916–25. https://doi.org/10.1016/j.jcf.2023.05.013.
[19] Tepper LA, Utens EMWJ, Caudri D, et al. Impact of bronchiectasis and trapped air on quality of life and exacerbations in cystic fibrosis. Europ Respirat J 2013;42:371–9. https://doi.org/10.1183/09031936.00137612.
[20] Wijker NE, Vidmar S, Grimwood K, et al. Early markers of cystic fibrosis structural lung disease: follow-up of the ACFBAL cohort. Eur Respir J 2020;55. https://doi.org/10.1183/13993003.01694-2019.
[21] Ramsey KA, Rosenow T, Turkovic L, et al. Lung clearance index and structural lung disease on computed tomography in early cystic fibrosis. Am J Respir Crit Care Med 2016;193:60–7. https://doi.org/10.1164/rccm.201507-1409OC.
[22] Kuo W, Soffers T, Andrinopoulou E-R, et al. Quantitative assessment of airway dimensions in young children with cystic fibrosis lung disease using chest computed tomography. Pediatr Pulmonol 2017;52:1414–23. https://doi.org/10.1002/ppul.23787.
[23] Tiddens H awm, Van Der Veer T, Charbonnier J-P, et al. Fully automated Mucus plug quantification on chest CTs and its correlation with all-cause mortality. A101. FULL METAL JACKET TARGETING COPD AND CHRONIC AIRWAYS DISEASE. American Thoracic Society; 2024. A2778–A2778.
[24] Mall MA. Unplugging mucus in cystic fibrosis and chronic obstructive pulmonary disease. Annals ATS 2016;13:S177–85. https://doi.org/10.1513/AnnalsATS.201509-641KV.
[25] Ramsey KA, Chen ACH, Radicioni G, et al. Airway mucus hyperconcentration in non–Cystic fibrosis bronchiectasis. Am J Respir Crit Care Med 2020;201:661–70. https://doi.org/10.1164/rccm.201906-1219OC.
[26] Mikami Y, Grubb BR, Rogers TD, et al. Chronic airway epithelial hypoxia exacerbates injury in muco-obstructive lung disease through mucus hyperconcentration. Sci Transl Med 2023;15:eabo7728. https://doi.org/10.1126/scitranslmed.abo7728.
[27] Aquino SL, Gamsu G, Webb WR, ST Kee. Tree-in-bud pattern: frequency and significance on thin section CT. J Comput Assist Tomogr 1996;20:594.
[28] Pieters ALP, van der Veer T, Meerburg JJ, et al. Structural lung disease and clinical phenotype in bronchiectasis patients: the EMBARC CT study. Am J Respir Crit Care Med 2024;210:87–96. https://doi.org/10.1164/rccm.202311-2109OC.

[29] Hoang-Thi T-N, Vakalopoulou M, Christodoulidis S, et al. Deep learning for lung disease segmentation on CT: which reconstruction kernel should be used? Diagn Interv Imaging 2021;102:691–5. https://doi.org/10.1016/j.diii.2021.10.001.
[30] Bakker JT, Klooster K, Wisselink HJ, et al. Effect of chest computed tomography kernel use on emphysema score in severe chronic obstructive pulmonary disease patients evaluated for lung volume reduction. Respiration 2023;102:164–72. https://doi.org/10.1159/000528628.
[31] de Jong PA, A Lindblad, Rubin L, et al. Progression of lung disease on computed tomography and pulmonary function tests in children and adults with cystic fibrosis. Thorax 2006;61:80–5. https://doi.org/10.1136/thx.2005.045146.