Comparison of Artificial Intelligence-Based Fully Automatic Chest CT Emphysema Quantification to Pulmonary Function Testing

Fischer, Andreas M.; Varga-Szemes, Akos; van Assen, Marly; Griffith, L. Parkwood; Sahbaee, Pooyan; Sperl, Jonathan I.; Nance, John W.; Schoepf, U. Joseph

Published in:
American Journal of Roentgenology

DOI:
10.2214/AJR.19.21572

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2020

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment.

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
Comparison of Artificial Intelligence–Based Fully Automatic Chest CT Emphysema Quantification to Pulmonary Function Testing

Andreas M. Fischer1,2
Akos Varga-Szemes1
Marly van Assen1,3
L. Parkwood Griffith1
Pooyan Sahbaee4
Jonathan I. Sperl5
John W. Nance6
U. Joseph Schoepf1

OBJECTIVE. The purpose of this study was to evaluate an artificial intelligence (AI)-based prototype algorithm for fully automated quantification of emphysema on chest CT compared with pulmonary function testing (spirometry).

MATERIALS AND METHODS. A total of 141 patients (72 women, mean age ± SD of 66.46 ± 9.7 years [range, 23–86 years]; 69 men, mean age of 66.72 ± 11.4 years [range, 27–91 years]) who underwent both chest CT acquisition and spirometry within 6 months were retrospectively included. The spirometry-based Tiffeneau index (TI; calculated as the ratio of forced expiratory volume in the first second to forced vital capacity) was used to measure emphysema severity; a value less than 0.7 was considered to indicate airway obstruction. Segmentation of the lung based on two different reconstruction methods was carried out by using a deep convolution image-to-image network. This multilayer convolutional neural network was combined with multilevel feature chaining and depth monitoring. To discriminate the output of the network from ground truth, an adversarial network was used during training. Emphysema was quantified using spatial filtering and attenuation-based thresholds. Emphysema quantification and TI were compared using the Spearman correlation coefficient.

RESULTS. The mean TI for all patients was 0.57 ± 0.13. The mean percentages of emphysema using reconstruction methods 1 and 2 were 9.96% ± 11.87% and 8.04% ± 10.32%, respectively. AI-based emphysema quantification showed very strong correlation with TI (reconstruction method 1, ρ = –0.86; reconstruction method 2, ρ = –0.85; both ρ < 0.0001), indicating that AI-based emphysema quantification meaningfully reflects clinical pulmonary physiology.

CONCLUSION. AI-based, fully automated emphysema quantification shows good correlation with TI, potentially contributing to an image-based diagnosis and quantification of emphysema severity.

Chronic obstructive pulmonary disease (COPD) is one of the most common diseases worldwide in terms of both morbidity and mortality [1–3]. Emphysema is characterized by the permanent overinflation of the terminal bronchioles and the destruction of bronchial walls and is caused by smoking tobacco in most cases occurring in Western countries [4–6]. Diagnosis of COPD is usually based on pulmonary function testing with spirometric measurements of forced expiratory volume in the first second (FEV1), forced vital capacity (FVC), and the Tiffeneau index (TI; FEV1/FVC) defined by the Global Initiative for Chronic Obstructive Lung Disease (GOLD) [7, 8]. In addition, the pathologic change of emphysema can be evaluated with quantitative CT (QCT) by measuring the low-attenuation volume (LAV) of the lung [9]. Although chest CT has not been included in the diagnostic workup recommended by the GOLD for patients with COPD, CT provides valuable supplementary information such as bronchial wall thickening, emphysema morphology, severity, and air trapping that directly contribute to airflow limitations and are associated with increased mortality rates [7, 8, 10–14]. In addition, other diseases that can clinically resemble COPD, such as pulmonary fibrosis, can only be differentiated by advanced imaging [15].

Several CT-based methods are commonly used in clinical investigations for the assessment of emphysema, including visual, semi-quantitative, and quantitative techniques [16–22]. All require strict acquisition parameters, need manual input, are time consuming, and are subject to increased interobserver and interpatient variability, especially...
tion consisted of 141 patients (Fig. 1). Thus, the final study population included age, sex, body mass index (BMI), race, smoking history, α-1-antitrypsin-deficiency, and current emphysema medication. Patients’ clinical reports were reviewed to collect information regarding emphysema diagnosis.

Study Protocol
Lung function test—All patients underwent pulmonary function testing (Vmax, CareFusion) including measurements of FVC, FEV1, and Tiffeneau index. TI. TI was used to measure emphysema severity. Excluding TI, all values are given as volumes and were calculated according to recommendations of the American Thoracic Society and European Respiratory Society [25] or Global Lung Function Initiative equations [26]. A TI below 0.7 was considered to represent airway obstruction.

Imaging protocol—CT studies were performed on three different systems (Somatom Definition Flash, Somatom Force, or Somatom Emotion, Siemens Healthineers). All studies were acquired during breath-hold at end inspiration and without administration of contrast media. The following scan parameters were used for the dual-source CT scanners: for second-generation scanners, collimation, 128 × 0.6 mm; tube voltage, 120 kV; and tube current, 75 mA; for third-generation scanners, collimation, 192 × 0.6 mm; tube voltage, 120 kV; and tube current, 80 mA. For the single-source CT scanner, we used 16 × 0.6 mm detector collimation, a tube voltage of 120 kV, and a tube current of 80 mA. To test the influence of various reconstruction methods (kernels) on AI performance, two reconstructions were analyzed in each of the 141 patients. Reconstruction method 1 used a section thickness of 1.5 mm with a lung kernel (filtered back-projection body kernel B60s, sharp, standard mode). Reconstruction method 2 used a section thickness of 1.5 mm with a soft-tissue kernel (filtered back-projection body kernel B31s, medium-smooth, standard mode). The mean CT dose index was 11.2 ± 23.6 mGy and the mean dose-length product was 302.8 ± 153.1 mGy · cm.

Image analysis—The segmentation of the lung was carried out using a deep image-to-image network (DI2IN), which is a multilayer convolutional neural network [27]. As a preprocessing step of the segmentation, the bronchial bifurcation landmark is detected in the 3D CT data using deep reinforcement learning [28]. If the landmark had not been found, the algorithm could not have been executed, but for all data used in this study the required landmark was found. The segmentation task is defined as the voxelwise classification. The algorithm takes the entire 3D CT volumes as input and outputs the probability maps that indicate how likely voxels are to belong to each lung lobe. The probability maps are provided with a threshold value. As shown in Figure 2, the main structure of DI2IN is designed using a symmetric way as a convolutional encoder-decoder. All blocks in DI2IN consist of 3D convolutional and bilinear upscaling layers. Figure 3 illustrates the training process, which used an adversarial training scheme. The batch normalization and the Leaky rectified linear unit were adopted in all convolutional layers for proper gradient back-propagation. The underlying network structure and the adversarial training strategy was identical to the one described by Yang et al. [27] in the context of liver segmentation.

The algorithm has been trained and tested on more than 10,000 CT datasets (approximately 5500 datasets were used for training). About 25% of the datasets were obtained from patients with emphysema. Other lung-related conditions represented were lung cancer, intestinal lung disease, consolidations, masses, pneumothorax, asthma, pneumonia, chronic bronchitis, and tuberculosis. The data originated from over 20 clinical sites in
the United States and Europe and were acquired with systems from three major CT vendors [27]. The data were annotated using in-house software. On the test data (approximately 4500 datasets), the mean Dice coefficients for the individual lung lobes ranged between 0.95 and 0.98 with a SD of 0.07 or less.

After AI-based lung segmentation, each CT scan was visually checked for correctness by a resident in radiology (2 years of clinical practice) and controlled by a radiologist (5 years of clinical experience after board certification). The airways in the pulmonary parenchyma were sufficiently excluded so that they did not erroneously contribute to the emphysema percentage.

Emphysema was quantified using spatial filtering and the commonly used threshold of –950 HU [16, 29, 30]. In a subset of 20 patients, the AI-based evaluation was repeated to evaluate test-retest reliability.

Statistical Analysis
Statistical analyses were conducted using SPSS software (version 23, IBM). Continuous variables are represented as mean ± SD or median (interquartile range) depending on how distribution was tested with the Shapiro-Wilk test. Categoric data are presented as absolute frequencies and proportions. Comparisons were made for all variables with each image reconstruction. Correlations between the TI and the emphysema quantification were analyzed with the Spearman correlation coefficient (\( \rho \)). Patients were categorized into emphysema-positive (TI < 0.7) and emphysema-negative groups (TI ≥ 0.7). Differences between these groups and differences between the two image types were compared using an independent \( t \) test. Test-retest reliability was assessed using an intraclass correlation coefficient (ICC). A \( p \) value less than 0.05 was considered statistically significant.

Results
Our population consisted of 69 men (49%) with a mean age of 66.72 ± 11.4 years (range, 27–91 years) and 72 women with a mean age of 66.46 ± 9.7 years (range, 23–86 years). Table 1 provides further demographic details, and Figure 4 shows representative images of patients with and without emphysema.

Lung Function Parameters
In the overall population, mean FEV\(_1\), FVC, and TI were 1.67 ± 0.69 L, 2.95 ± 0.92 L, and 0.57 ± 0.13, respectively. Pathologic TI was found in 104 patients (TI = 0.50 ± 0.13), whereas 37 patients showed physiologic TI of 0.74 ± 0.04. Table 2 provides a detailed overview of the pulmonary function testing results.

| TABLE 1: Clinical Characteristics and Demographics of Study Participants |
|----------------|-----------------|---------------|---------------|
| Characteristic | All Patients | TI < 0.7 | TI ≥ 0.7 |
| Total no. | 141 | 104 | 37 |
| Sex | | | |
| Female | 72 (51) | 51 (49) | 21 (57) |
| Male | 69 (49) | 53 (51) | 16 (43) |
| Mean age ± SD (y) | 66.59 ± 9.13 | 67.69 ± 8.56 | 63.51 ± 12.8 |
| Mean body mass index ± SDa | 26.22 ± 6.26 | 25.46 ± 5.98 | 28.35 ± 6.27 |
| Smoker | 124 (88) | 104 (100) | 20 (54) |
| \( \alpha \)-1-antitrypsin deficiency | 1 (1) | 1 (1) | 0 |
| Medication | | | |
| Salbutamol | 65 (46) | 64 (62) | 1 (3) |
| Formoterol | 27 (19) | 26 (25) | 1 (3) |
| Vilaanterol | 3 (2) | 3 (3) | 0 |
| Fluticasone | 18 (13) | 17 (16) | 1 (3) |
| Budesonide | 34 (24) | 34 (33) | 0 |
| Tiotropium bromide | 23 (16) | 23 (22) | 0 |
| Ipratropium bromide | 11 (8) | 11 (11) | 0 |
| Theophylline | 3 (2) | 3 (3) | 0 |
| Prednisolone | 3 (2) | 2 (2) | 1 (3) |

Note—Values in parentheses are percentages. TI = Tiffeneau index.
aWeight in kilograms divided by the square of height in meters.

| TABLE 2: Lung Function Results of Study Participants |
|----------------|-----------------|---------------|---------------|
| Result | All Patients \( (n = 141) \) | TI < 0.7 \( (n = 104) \) | TI ≥ 0.7 \( (n = 37) \) |
| FEV\(_1\) (L) | 1.67 ± 0.69 | 1.49 ± 0.58 | 2.16 ± 0.62 |
| FVC (L) | 2.95 ± 0.92 | 2.96 ± 1.02 | 2.92 ± 0.88 |
| TI (%) | 0.57 ± 0.13 | 0.50 ± 0.13 | 0.74 ± 0.04 |
| Emphysema percentage | | | |
| Reconstruction method 1a | 9.96 ± 11.87 | 13.79 ± 12.60 | 1.12 ± 0.95 |
| Reconstruction method 2a | 8.04 ± 10.32 | 11.14 ± 10.98 | 1.08 ± 1.70 |

Note—Values are mean ± SD. TI = Tiffeneau index, FEV\(_1\) = forced expiratory volume in 1 second, FVC = forced vital capacity.
aSection thickness of 1.5 mm with a lung kernel (filtered back-projection body kernel B60s, sharp, standard mode).
bSection thickness of 1.5 mm with a soft-tissue kernel (filtered back-projection body kernel B31s, medium-smooth, standard mode).
Emphysema Values

In the context of reconstruction method 1, 104 patients with a pathologic TI showed a mean emphysema percentage of 13.79% ± 12.60%, whereas 37 patients with a physiologic TI showed a mean emphysema percentage of 1.12% ± 0.95%. With reconstruction method 2, 104 patients with a pathologic TI showed a mean emphysema percentage of 11.14% ± 10.98%, and 37 patients with a physiologic TI showed a mean emphysema percentage of 1.08% ± 1.70%. Table 2 gives an overview of the emphysema quantification data for our patients.

Correlations Between Tiffeneau Index and Emphysema

AI-based emphysema quantification showed a very strong correlation with TI, presenting correlation coefficients of $\rho = -0.86$ ($p < 0.0001$) and $\rho = -0.85$ ($p < 0.0001$) when compared with reconstruction methods 1 and 2, respectively. Thus, the AI algorithm tested here was able to meaningfully predict the decrease in lung function that is associated with a higher percentage of the lung being affected by emphysema. Furthermore, we could confirm that the two reconstruction methods yielded nearly identical correlations between emphysema quantification and TI, so the AI-based lung segmentation should apply equally for soft-tissue and lung reconstruction settings. Table 3 gives an overview of the correlation between the TI and emphysema percentage. Figure 5 shows the scatterplot of the relationship between TI and LAV of both reconstruction methods.

Test-Retest Reliability

Calculation of LAV by both reconstruction methods showed high test-retest reliability (for reconstruction method 1, ICC = 1.0 [0.9–1.0]; for reconstruction method 2, ICC = 1.0 [1.0–1.0]).

Table 3: Correlation of FEV\textsubscript{1} and Tiffeneau Index for Emphysema Quantification

<table>
<thead>
<tr>
<th>Emphysema Percentage</th>
<th>FEV\textsubscript{1}</th>
<th>Tiffeneau Index</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\rho$</td>
<td>$p$</td>
</tr>
<tr>
<td>Reconstruction method 1\textsuperscript{a}</td>
<td>$-0.57$</td>
<td>$&lt;0.0001$</td>
</tr>
<tr>
<td>Reconstruction method 2\textsuperscript{b}</td>
<td>$-0.56$</td>
<td>$&lt;0.0001$</td>
</tr>
</tbody>
</table>

Note—FEV\textsubscript{1} = forced expiratory volume in 1 second.

\textsuperscript{a}Section thickness of 1.5 mm with a lung kernel (filtered back-projection body kernel B60s, sharp, standard mode).

\textsuperscript{b}Section thickness of 1.5 mm with a soft-tissue kernel (filtered back-projection body kernel B31s, medium-smooth, standard mode).
quantitative CT analysis of the lung to deter-
mine the extent of pulmonary emphysema. Our results suggest that AI can be success-
fully used to measure the increase in LAV
that relates to a decrease in TI. To deter-
mine whether the AI-based segmentation of the lung boundaries performs independ-
ently of the reconstruction method and that the determined LAVs show similar correlation with lung function parameters, two different reconstruction methods were analyzed for each patient and showed nearly identical results. However, we did find that the B60s ker-
nel detected, on average, a higher proportion of emphysema than did the B30s kernel. Gi-
erada et al. [31] reported a higher percentage of emphysema for the B60f kernel than for the B30f kernel. However, a possible effect of the different kernels must also be taken into account here. Gierada et al. used B60f and B30f kernels, whereas our study used B60s and B31s kernels, both of which have an edge enhancement feature. Whether differ-
ent parameters also affect the correlation of the dynamic lung function parameters is out-
side the scope of this article. We were able to show that the two reconstruction methods (using B60s and B31s kernels) in this study had an insignificant effect on correlation with the spirometric data.

Emphysema quantification has been a top-
ic of research for many years because of the labor-intensive process it requires and the subjectivity of visually quantifying emphy-
sema. Perhaps because of the challenges in providing truly quantitative data, GOLD has not mandated chest CT as a step in the diagnostic workup of patients suspected of hav-
ing COPD [8]. However, compared with spi-
rometry, CT provides additional insight into pathologic changes that directly contribute to airway obstruction such as parenchymal de-
struction, thickening of the bronchial wall, and increased air retention by air trapping
[10, 13, 32, 33]. Moreover, interstitial lung disease and pulmonary fibrosis, which could symptoms overlapping with those of COPD, can also be distinguished by imaging [15].

Bankier et al. [21] found that QCT mea-
urements correlated better with macro-
scopic emphysema measurements than with visual evaluation. Regarding threshold selec-
tion, Müller et al. [17] examined the respec-
tive segments of 62 patients from high-reso-
lution CT before partial lung resection using a threshold of –950 HU. Additionally, Mada-
ni et al. [30] reported on 80 patients under-
going lung resection and found that the best correlation between microscopic and macro-
sopic pulmonary emphysema occurs using a density mask method at –960 HU or –970
HU. However, Schroeder et al. [32] noted that milder emphysematous changes could be missed at a threshold of –960 HU or –970
HU. Therefore, they chose to use –950 HU
and reported a correlation between the LAV in inspiratory chest CT and the TI of

r = –0.76 in over 4000 patients.

Some studies have compared QCT and
TI in smaller populations (14–26 patients) and reported similar results, with correla-
tions ranging between 0.75 and 0.79 [34, 35]. However, both of these studies were based on a semiautomatic evaluation system. This time-consuming approach required man-
ual readjustments to fit the outer limits of the lung during image processing and were cor-
rected according to the actual anatomy. The AI-based fully automatic lung segmentation captures a correct and reproducible anatomy of the respective lung, which ensures the nec-

dessary prerequisite that the LAV of the lung tissue is also fully captured and thus, in con-
trast to the semiautomatic evaluation, ensures the evidence of the results.

In addition to the considerable labor inten-
sity with manual or semiautomatic evaluation, there is an increase in interobserver variabil-
ity [24, 36–38]. Despite good correlations between lung function values and extent of pulmonary emphysema, the evaluation strat-

ey in today’s clinical routine cannot correct for subjectivity and time commitment with an economy-driven day-to-day workflow
[24]. In contrast, after our AI prototype was trained, the fully automatic AI-based eval-
uation properly recorded all lung bound-
aries and resulted in very strong correlations with TI and emphysema quantification, re-
gardless of the reconstruction method used. These observations suggest the reproducibil-
ity of our quantitative imaging results along with enhanced objectivity, and evaluations were completed in a mean of 60 seconds. Although AI is considered one of the most promising areas in medical imaging and emphysema quantification should be an intui-
tive arena of application, we are not aware of any comparable studies on AI-based emphysema quantification [39]. The ability of an AI-based system to analyze high-dimen-
sional image data and generate valuable clinical information without input from human observers should thus hold considerable po-
tential for granular quantitative imaging ap-
proaches to provide patients with accurate diagnoses and steer treatment.
Limitations
This retrospective study is subject to several limitations that should be considered in the interpretation of our results. First, the CT scan did not include spirometric triggering, so we could not completely ensure that imaging studies were obtained during maximum inspiration. Furthermore, only spirometry data were collected and compared with QCT. Because of the correlation of the static LAV parameters with the dynamic lung function parameter of spirometry, nearly 20% of the patients were excluded from the study because we wanted to ensure that the obstruction in the lung function was mainly due to emphysema and not to a compression effect from a condition that influences exhalation, such as an auxiliary tumor. Therefore investigating the correlation between static body–plethysmographic lung function parameters and static QCT of emphysema would be an interesting focus for future studies that could also include patients with cystic lung diseases, bronchiectasis, or hilar masses. Lastly, further work is needed to establish QCT analysis as an imaging biomarker for patients with COPD, providing additional insights into the clinical picture of COPD and potentially working as a common diagnostic tool alongside pulmonary function tests.

In conclusion, we found that the AI-based algorithm introduced here was able to quantify emphysema rapidly and without input by human observers, with very strong correlation to man observers, with very strong correlation to clinical pulmonary function testing parameters.

Clinical Outlook
AI-based fully automated emphysema quantification shows good correlation with TI, potentially contributing to an image-based diagnosis and quantification of severity in patients with emphysema, along with their long-term follow-up. Further investigation is needed to establish quantified AI-based emphysema analysis as a potential biomarker for patients with COPD, thus improving diagnostic performance for a specific outcome and maximizing information retrieval to better understand the causes of disease. Thus, AI-based pulmonary emphysema diagnostics could contribute to complementary phenotyping as part of an imaging biomarker for patients with COPD to shape the individual therapy of patients as a common diagnostic tool, in combination with pulmonary function tests.

References
Quantifying Emphysema With AI and Pulmonary Function Testing

Pattern Anal Mach Intell 2019; 41:176–189


36. Dournes G, Montaudon M, Berger P, Laurent F. In vivo computed tomography as a research tool to investigate asthma and COPD: where do we stand? J Allergy (Cairo) 2012; 2012:972479


FOR YOUR INFORMATION

ARRS is accredited by the Accreditation Council for Continuing Medical Education (ACCME) to provide continuing medical education activities for physicians.

The ARRS designates this journal-based CME activity for a maximum of 1.00 AMA PRA Category 1 Credits™ and 1.00 American Board of Radiology®, MOC Part II, Self-Assessment CME (SA-CME). Physicians should claim only the credit commensurate with the extent of their participation in the activity.

To access the article for credit, follow the prompts associated with the online version of this article.