U-Net-based Lung Thickness Map for Pixel-level Lung Volume Estimation of Chest X-rays

Preprint

Tina Dorosti*, Manuel Schultheiß*, Philipp Schmette, Jule Heuchert, Johannes Thalhammer, Florian Schaff, Thorsten Sellerer, Rafael Schick, Kirsten Taphorn, Korbinian Mechlem, Lorenz Birnbacher, Franz Pfeiffer, Daniela Pfeiffer

Key points:

1. Synthetic frontal chest X-ray radiographs and two-dimensional lung thickness maps were generated from CT reconstructions using a forward projection model based on physical processes to ensure applicability across different X-ray machines.

2. Pixel-level lung volumes were estimated for chest X-ray radiographs using predicted lung thickness maps by a U-Net.

3. Quantitative analysis of thickness maps and qualitative comparison of resulting total lung volume estimations showed promising results for synthetic and real radiographs.

Summary statement:

With CT-derived total lung volume as ground truth, strong correlations were measured between the U-Net predicted and ground truth lung volumes for real radiographs (Pearson’s $r = 0.908$, $P < 0.001$).

Abbreviations: Confidence Interval (CI), Convolutional Neural Network (CNN), Chronic Obstructive Pulmonary Disease (COPD), Computed Tomography (CT), Hounsfield Units (HU), Mean Squared Error (MSE), Mean Absolute Error (MAE), Pulmonary Embolism (PE), Total Lung Volume (TLV)

1 Chair of Biomedical Physics, Department of Physics, School of Natural Sciences, Technical University of Munich, 85748 Garching, Germany
2 Munich Institute of Biomedical Engineering, Technical University of Munich, 85748 Garching, Germany
3 Department of Diagnostic and Interventional Radiology, School of Medicine and Health, Klinikum rechts der Isar, Technical University of Munich, 81675 Munich, Germany
4 Institute for Advanced Study, Technical University of Munich, 85748 Garching, Germany

* Equal contribution
Abstract

**Purpose:** We aimed to estimate the total lung volume (TLV) from real and synthetic frontal X-ray radiographs on a pixel level using lung thickness maps generated by a U-Net.

**Methods:** 5,959 thorax X-ray computed tomography (CT) scans were retrieved from two publicly available datasets of the lung nodule analysis 2016 (n=656) and the RSNA pulmonary embolism detection challenge 2020 (n=5,303). Additionally, thorax CT scans from 72 subjects (33 healthy: 20 men, mean age [range] = 62.4 [34, 80]; 39 suffering from chronic obstructive pulmonary disease: 25 men, mean age [range] = 69.0 [47, 91]) were retrospectively selected (10.2018-12.2019) from our in-house dataset such that for each subject, a frontal chest X-ray radiograph no older than seven days was available. All CT scans and their corresponding lung segmentation were forward projected using a simulated X-ray spectrum to generate synthetic radiographs and lung thickness maps, respectively. A U-Net model was trained and tested on synthetic radiographs from the public datasets to predict lung thickness maps and consequently estimate TLV. Model performance was further assessed by evaluating the TLV estimations for the in-house synthetic and real radiograph pairs using Pearson correlation coefficient (r) and significance testing.

**Results:** Strong correlations were measured between the predicted and CT-derived ground truth TLV values for test data from synthetic \(n_{\text{public}}=1,191\), \(r=0.987, P < 0.001\); \(n_{\text{in-house}}=72\), \(r=0.973, P < 0.001\) and real radiographs \(n=72\), \(r=0.908, P < 0.001\).

**Conclusion:** TLV from U-Net-generated pixel-level lung thickness maps were successfully estimated for synthetic and real radiographs.

**Keywords:** Frontal chest X-ray radiographs, Lung thickness map, Pixel-level, Total lung volume, U-Net
1 Introduction

Numerous lung conditions, including infectious diseases, interstitial lung diseases, and chronic obstructive pulmonary disease (COPD), affect lung function by either reducing or increasing the air volume within the lungs [1–4]. Total lung volume (TLV) is a critical metric for evaluating the severity, progression, and treatment response in cases of restrictive lung diseases. Pulmonary function tests, such as spirometry and body plethysmography, are commonly conducted to obtain the total lung capacity. Imaging techniques offer an alternative and are of particular interest for patients unable to perform pulmonary function tests or when the lung has already been imaged [5].

Traditionally, image-based methods for estimating lung volume use the air volume at maximum inspiration in lateral and frontal chest X-ray radiographs. Hurtado et al. manually calculated the overall lung area and multiplied it by the PA diameter [2]. Pierce et al. used shape information to gain a more accurate estimate of the volume [6]. Alternatively, X-ray computed tomography (CT) allows for an accurate estimation of the lung volume by multiplying the area by the thickness of each slice over the entire volume [5].

More recently, deep-learning-based methods have shown promising results [7-13]: With a combination of pulmonary function test results, CT-derived data, as well as chest X-ray radiographs, deep-learning models can be trained to estimate TLV [7-8, 10, 12]. Boulogne et al. [11] use a deep learning architecture for CT data to predict spirometry and diffusion capacity of carbon monoxide estimates at the patient and lobe levels. Marsh explored the generation of relative lung thickness maps for CT-simulated radiographs [13]. To the best of our knowledge, there are no deep-learning models available for pixel-level lung volume estimation of real chest X-ray radiographs.

Pixel-level thickness maps can provide the location and shape information of healthy and dysfunctional areas across the lung. Such maps can be used to accurately measure volumetric changes in pulmonary tissue, improving risk assessment for pulmonary resection surgery [11, 14]. Furthermore, pixel-level lung volume estimation could be used in novel imaging modalities such as
X-ray dark-field imaging, where dividing the measured signal by the lung thickness improves the image contrast [15-16].

In this exploratory study, we aim to estimate the lung volume on a pixel level using lung thickness maps generated by a U-Net. We simulate synthetic frontal chest X-ray radiographs and two-dimensional lung thickness maps from CT data and volumetric lung segmentations using a forward projection model based on physical processes. This approach ensures applicability across different X-ray machines. We assess the model’s performance using both synthetic and real radiographs.

2 Methods

The code is available at: https://github.com/tidorosti/lungThickness

2.1 Datasets

For synthetic frontal chest X-ray radiograph generation, thorax CT scans were retrieved from two publicly available datasets of the lung nodule analysis 2016 (Luna16) challenge [17] (n=888) and the RSNA Pulmonary Embolism (PE) detection challenge 2020 [18] (n=7,122). For comparability reasons, only CT scans acquired with 120 kVp were considered, resulting in a total of 656 CT scans from the Luna16 dataset and 5,303 CTs from the PE dataset. Thorax CT scans from 72 subjects (33 healthy, 39 suffering from COPD) were selected from our in-house dataset such that for each subject, a medically indicated frontal chest radiograph no older than seven days was available. We received approval from our institutional review board, and the requirement for written informed consent was waived, as data was analyzed anonymously and retrospectively (10.2018-12.2019). The subjects in the in-house dataset were previously reported on as part of a prior publication focused on X-ray dark-field chest imaging for assessment of emphysema in patients with COPD [16]. In the current study, we utilize the retrospectively selected CT data and corresponding chest X-ray attenuation images from these subjects to assess the performance of the proposed U-Net on pixel-level lung thickness estimation. Subject demographics for the in-house data are provided in Table 1.
The complete data selection flowchart for both the public and the in-house datasets is shown in Fig. 1.

2.2 Data Preparation

Before the synthetic X-ray radiographs were generated from the CT data, the patient table was segmented and removed from the volume. To do so, a binary threshold at -775 Hounsfield unit (HU) was set to separate the air and the body for each CT slice image. An opening filter removed thin lines from partial volume effects between the table and the volume. The torso of the body was selected by finding the biggest connected object with a connected component algorithm. Python’s ndimage library from the SciPy package (version 1.10.1) was used [19].

The lung segmentations for the PE and the Luna16 CT data were obtained with the lungmask package (version 0.2.20) [20]. For the in-house data, the automated lung segmentation from CT, as well as ground truth TLV estimations, were supplied by the CT scanner manufacturer (IntelliSpace Portal, Philips Medical Systems, Hamburg, Germany).

2.3 Synthetic Radiograph and Thickness Map Generation

The following method was applied to generate an X-ray spectrum for simulating synthetic radiographs and their corresponding thickness maps via forward projection. An overview of the simulation process is shown in Fig. 2A.

2.3.1 X-ray Spectrum Simulation

Certain standard parameters of X-ray systems were set to simulate realistic synthetic radiographs from CT, namely, the tube voltage (70 kVp) and detector properties, such as detector thickness (0.6 mm) and scintillator material (cesium iodide, density=3.38 g/cm³). The tube voltage was selected to match that of the setup used to measure the real radiographs at our clinics. The X-ray spectrum was simulated with a semi-empirical model for X-ray transmission using SpekPy (version 2.0.1) [21-23].
2.3.2 Material Segmentation

To attribute correct attenuation properties to the different tissue types in the human thorax, the CT scans were segmented into three basis material volumes by applying the following voxel-level thresholds: (-200 – 0) HU for adipose tissue, (0 – 240) HU for soft tissue, and values above 240 HU for bone. These values fell within the ranges described by Buzug et al. [24] and were slightly adjusted to avoid overlapping or missing HU ranges.

To account for differences in material densities, \( \rho_i \), which are reflected in slight variations in the measured HU values, adjusted attenuation values, \( \mu_i'(E) \), were calculated for the basis material \( i \) in each voxel [25]:

\[
\mu_i'(E) = \frac{\mu_i(E)}{\rho_i} \cdot \rho_i'.
\]  

(1)

Here, \( \frac{\mu_i}{\rho_i}(E) \) is the tabulated energy-dependent mass attenuation coefficient of basis material \( i \), adjusted with the thresholded voxel-wise mass density \( \rho_i'. \) Considering the definition of the HU value:

\[
HU = \frac{\mu - \mu_{water}}{\mu_{water}} \cdot 1000,
\]  

(2)

the latter is derived from the measured HU value of material \( i \) in a given voxel, \( HU_i \), as follows:

\[
\rho_i' = \frac{HU_i}{1000} \cdot \frac{\mu_{water}(E_{CT}) + \mu_{water}(E_{CT})}{\frac{\mu_i}{\rho_i}(E_{CT})}.
\]

(3)

The tabulated attenuation coefficients of water, \( \mu_{water}(E_{CT}) \), and the mass attenuation coefficient of basis material \( i \), \( \frac{\mu_i}{\rho_i}(E_{CT}) \), were retrieved from the NIST database [26] using the xraylib framework (version 4.1.3) [27]. These values were obtained for \( E_{CT} = 70 \) keV, which is the estimated mean energy for the 120 kVp spectrum of the origin CT scanner.
2.3.3 Forward Projection

For each CT scan, the density volumes and the corresponding lung segmentation volume were forward-projected to obtain the synthetic radiograph projections and the matching 2-dimensional ground-truth lung-thickness map. A cone-beam projector was implemented with the Astra toolbox (version 2.1.3) [28] using the following parameters: detector size = 512x512 pixels, source-to-sample distance = 2,060 mm, and sample-to-detector distance = 375 mm. 10 projections were created for each sample by taking measurements from -10° to 10° at 2° steps.

The intensity value \( I \) of each pixel in the synthetic radiograph was calculated with

\[
I = \sum_{E=1}^{K} \Phi(E) \cdot \exp \left( \sum_{i=1}^{N} \frac{\mu_i(E) \cdot \rho_i}{\rho_i} \right),
\]

given the energy-dependent number of photons \( \Phi(E) \), a kilo-voltage peak \( K \), and \( N \) basis materials.

Lastly, the negative logarithm of the intensity, normalized with respect to the flat-field image, was taken for a conventional depiction of the radiograph.

2.4 U-Net Implementation

A U-Net architecture [29] with six layers of depth was utilized for input images downsampled to 256x256 pixels. The model had 3,718,433 parameters and was trained from scratch, with randomly initialized weights, for 120 epochs with a learning rate of \( 10^{-4} \) and a batch size of 8. A GeForce RTX 3090 NVIDIA graphics processing unit with 24 GB of VRAM was utilized. Using Keras [30] and TensorFlow [31] (versions 2.11.0), the U-Net was trained with the mean squared error (MSE) loss and adaptive moment estimation optimizer. The final model was selected from the epoch with the best validation loss. The detailed model architecture is depicted in Fig. 2B.

Synthetic data from the public datasets were shuffled using Python’s (version 3.8.10) built-in function and divided on a train-validation-test split with a 60:20:20 ratio. The data in the train and validation splits were balanced between the two public datasets as follows: For every CT scan, all 10
simulated projections were utilized for Luna16 data, but only the central projection at 0° from PE data was considered. In total, there were 7,302 images for training ($n_{\text{Luna16}} = 4,120, n_{\text{PE}} = 3,182$) and 2,191 images for validating ($n_{\text{Luna16}} = 1,130, n_{\text{PE}} = 1,061$). Additional data augmentation was implemented with Keras’ data generator and applied to the simulated radiographs by randomly shifting the images vertically and horizontally within a range of 0.2 relative to the image dimension and rotating the images within a range of 10°. The model was tested on 1,191 synthetic radiographs generated from public data ($n_{\text{Luna16}} = 131, n_{\text{PE}} = 1,060$), and 72 ($n_{\text{Healthy}} = 33, n_{\text{COPD}} = 39$) image pairs corresponding to real and synthetic radiograph pairs from the in-house data. All synthetic test data was retrieved from the central projections at 0°.

2.5 Evaluation Metrics and Statistical Analysis

Statistical normality for data distribution was confirmed with the Shapiro-Wilk, D’Agostino’s K-squared, and Anderson-Darling tests before calculating the 95% confidence intervals (CI) for mean TLV. Model performance was assessed with the MSE and mean absolute error (MAE) between the ground truth and predicted TLV for the synthetic and real test data. Pearson correlation coefficient calculation and significance testing were performed. The statistical analysis was done with SciPy (version 1.10.1) [19]. Linear fits were calculated with NumPy (version 1.24.3) [32].

3 Results

The results of testing the U-Net on 1,191 synthetic radiographs from public data, as well as 72 synthetic and real radiograph pairs from our in-house data (33 healthy subjects: 20 men, mean age [range] = 62.4 [34, 80]; 39 suffering from COPD: 25 men, mean age [range] = 69.0 [47, 91]), are provided in this section.

3.1 Synthetic Radiographs

Quantitative results of TLV estimation for the synthetic radiographs are presented in Table 2. A strong correlation exists between the predicted and the ground truth mean TLV values for all test
data. The null hypothesis \((r=0)\) was rejected at a significance level of 0.01 \((P < 0.001)\). Luna16 test data resulted in the highest correlation \((r = 0.993, P < 0.001)\) and the lowest error \((\text{MSE} = 0.094 \text{ L}, \text{MAE} = 0.230 \text{ L})\). The COPD data resulted in the lowest correlation \((r = 0.968, P < 0.001)\) and the highest error \((\text{MSE} = 0.282 \text{ L}, \text{MAE} = 0.379 \text{ L})\). The model tends to moderately underestimate the TLV.

Fig. 3 shows radiographs with their ground truth and predicted thickness maps for eight example cases from the synthetic public data. There are no considerable qualitative differences between the first four cases corresponding to the Luna16 data and the last four cases corresponding to the PE data. Examples of eight synthetic radiographs from the healthy and COPD in-house data, the corresponding predicted and ground truth lung thickness maps, and their absolute difference are provided in Fig. 4A-D. For both public and in-house data, the predicted thicknesses are somewhat smoother and less detailed than the ground truth thickness maps. Qualitative thickness map differences are more prominent for the in-house data.

### 3.2 Real Radiographs

TLV estimations for real radiographs are given in Table 3. Pearson correlation coefficients are slightly lower, and MSE and MAE values are higher for real radiographs in comparison to their synthetic counterparts. Nonetheless, predicted and ground truth TLV values for real radiographs are strongly correlated \((P < 0.001)\). The predicted TLVs for real radiographs are marginally underestimated in comparison to the CT-derived ground truths supplied by the manufacturer.

Real radiographs and corresponding predicted thickness maps for four healthy subjects and four subjects with COPD are shown in Fig. 4E-F. A qualitative comparison of predicted lung thickness maps for the synthetic and real radiograph pairs is provided in Fig. 4. In some cases, the thicknesses were underestimated for the real radiograph of the same subject in Fig. 4F when compared to the ground truth maps generated from CT lung masks in Fig. 4C. Patient posture differences such as the position of the arms, variations in inspiration cycle, as well as the time-gap between CT and
radiograph acquisitions must be considered when comparing real and synthetic radiograph pairs and their predicted thicknesses.

Scatter plots with linear fits of ground truth versus predicted TLV for individual subjects are shown for the public test data in Fig. 5A, the synthetic in-house data in Fig. 5B, and the real in-house data in Fig. 5C. For the in-house data, synthetic radiographs had a larger correlation coefficient ($r = 0.973, P < 0.001$) than their real radiograph pairs ($r = 0.908, P < 0.001$). The strongest correlation ($r = 0.987, P < 0.001$) and the best linear fit (slope = 1.01, offset = -0.023 L) were given by synthetic radiographs from the public datasets.

The model performance was separately assessed for four real radiographs that included foreign objects. These cases are depicted in Fig. 6, showing the real radiographs and the predicted thickness maps, two of which correspond to healthy subjects and two to subjects with COPD. In all cases, the strongly absorbing foreign object present in the scan causes changes to the intensity range in the radiograph, which in turn results in failed thickness map predictions by the model. These cases were excluded from the in-house data, as shown in Fig 1B, and were only investigated for the failure analysis in Fig 6.

4 Discussion

We trained a U-Net for pixel-level lung thickness map and TLV estimation using synthetic radiographs and applied the trained model to real radiographs. The results obtained from both quantitative and qualitative analysis of synthetic and real radiographs were promising: Taking CT-derived TLV values as ground truth, strong correlations were measured between the predicted and ground truth TLV values for synthetic ($r_{\text{Public}} = 0.987, P < 0.001; r_{\text{In-house}} = 0.973, P < 0.001$) and real radiographs ($r = 0.908, P < 0.001$).

Larger positive Pearson’s correlation coefficients and lower MAE values showed that the model was able to better predict TLV values for the synthetic public test data in comparison to the synthetic in-house test data. This behavior is explainable as the model train and validation data splits
were also drawn from the same public data distribution. Therefore, it is understandable that the model would have sub-par performance on the external synthetic data taken from our in-house data distribution.

We examined the robustness of our method for multiple lung pathologies, namely lung nodules, PE, and COPD, in comparison to healthy lungs. For both synthetic and real data, regardless of the pathology, our ground truth and predicted TLV results were within previously reported CT-derived mean TLV range for healthy lungs as described by Haas et al. [33] (n=302, mean ± standard deviation = 5.28±0.947 L) and Wisselink et al. [34] (n_femal=151, mean ± standard deviation = 4.7±0.9 L; n_male=139, mean ± standard deviation = 6.2±1.2 L). Our findings for the PE test set fell towards the lower end of this range, which can be attributed to the inclusion of acute and chronic PE types in the RSNA PE dataset [18], given subjects with acute PE experience a reduction of pulmonary function and capacity [35].

Compared to previous work on lung volume estimation with deep-learning models for real radiographs by Sogancioglu et al. [7] (MAE = 0.592 L, r = 0.855), our model achieved better results (MAE = 0.462 L, r = 0.908). Additionally, we are able to obtain not only the total lung volume but also a pixel-level lung thickness map. Since such a lung thickness map cannot be calculated from real radiographs due to the missing ground truth information, the training must be performed exclusively with simulated X-rays calculated from CT scans. Here, it should be emphasized that we are still able to test the model on real, non-simulated, radiographs and report promising results based on an evaluation with respect to the CT-derived total lung volume.

However, transferring knowledge from CT scans to real radiographs presents several inherent hurdles, mainly different patient postures and varying inspiration levels in CT data [7]. Additionally, neglected physical effects such as Compton scattering could also account for differences between real and synthetic radiographs. Yet, our results for synthetic radiographs indicate that with further consideration of these effects, even lower predicted TLV errors can be achieved for real radiographs.
The main limitation of this work was the rather small sample of real radiographs. We used unprocessed radiographs from our in-house setup to best match with the synthetically generated radiograph pairs and to avoid the need for vendor-specific post-processing algorithms, which are closed sources and unavailable from the imaging device manufacturers. Furthermore, the current workflow could benefit from additional improvements. For example, deep learning-based segmentation can be used for different tissue types instead of HU thresholding. Future research could investigate the additional use of lateral radiographs for thickness map estimation and explore other network architectures.

In conclusion, TLV is a critical metric for evaluating the severity, progression, and treatment response in various lung diseases, such as interstitial pulmonary fibrosis and COPD. Pixel-level lung thickness maps for real and synthetic radiographs can be generated with a U-Net trained on synthetic CT-derived radiographs for subsequent calculation of the TLV. Strong correlations are observed between the CT-derived ground truth and the U-Net predicted TLV values obtained from the generated thickness maps.

References

[1] Biselli P, Grossman PR, Kirkness JP, et al. The effect of increased lung volume in chronic obstructive pulmonary disease on upper airway obstruction during sleep. J Appl Physiol 2015;119:266-271, doi: 10.1152/japplphysiol.00455.2014

[2] Hurtado A, Fray WW, Kaltreider NL, Brooks WD. Studies of total pulmonary capacity and its subdivisions. v. normal values in female subjects. J Clin Invest 1934;13:169-191, doi: 10.1172/JCI100576

[3] van Noord JA, Clément J, Cauberghs M, Mertens I, Van de Woestijne KP, Demedts M. Total respiratory resistance and reactance in patients with diffuse interstitial lung disease. Eur Clin Respir J 1989;2:846–852.
[4] Huang Y, Tan C, Wu J, et al. Impact of coronavirus disease 2019 on pulmonary function in early convalescence phase. Respir Res 2020;21:163. doi:10.1186/s12931-020-01429-6

[5] Flesch JD, Dine C J. Lung volumes. Chest 2012;142:506-510, doi: 10.1378/chest.11-2964

[6] Pierce RJ, Brown DJ, Holmes M, Cumming G, Denison DM. Estimation of lung volumes from chest radiographs using shape information. Thorax 1979,34:726-734. doi:10.1136/thx.34.6.726

[7] Sogancioglu E, Murphy K, Th Scholten E, Boulogne LH, Prokop M, van Ginneken B. Automated estimation of total lung volume using chest radiographs and deep learning. Med Phys 2022;49:4466-4477. doi:10.1002/mp.15655

[8] Kim H, Jin KN, Yoo SJ, et al. Deep Learning for Estimating Lung Capacity on Chest Radiographs Predicts Survival in Idiopathic Pulmonary Fibrosis. Radiology 2023; 306:e220292. doi:10.1148/radiol.220292

[9] Beverin L, Topalovic M, Halilovic A, Desbordes P, Janssens W, De Vos M. Predicting total lung capacity from spirometry: a machine learning approach. Front Med (Lausanne) 2023;10:1174631. doi:10.3389/fmed.2023.1174631

[10] Wysoczanski A, Angelini ED, Sun y, et al. Multi-view cnn for Total lung volume inference on cardiac computed tomography. IEEE ISBI 2023;1-5. doi: 10.1109/ISBI53787.2023.10230821

[11] Boulogne LH, Charbonnier JP, Jacobs C, van der Heijden EHFM, van Ginneken B. Estimating lung function from computed tomography at the patient and lobe level using machine learning. Med Phys 2024;51:2834-2845. doi:10.1002/mp.16915

[12] Ghimire S, Subedi S. Estimating lung volume capacity from x-ray images using deep learning. Quantum Beam Sci 2024;8:11. doi: 10.3390/qubs8020011

[13] Marsh. 2D lung thickness estimation from chest x-rays using U-Net regression trained with digitally reconstructed radiographs. 2024. University of Dayton, Master's thesis. OhioLINK Electronic Theses and Dissertations Center,
[14] Sengul AT, Sahin B, Celenk C, Basoglu A. Postoperative lung volume change depending on the resected lobe. Thorac Cardiovasc Surg 2013;61:131-137. doi:10.1055/s-0032-1322625

[15] Pfeiffer F, Weitkamp T, Bunk O, David C. Phase retrieval and differential phase-contrast imaging with low-brilliance X-ray sources. Nature Phys 2006;2:258–261. doi:10.1038/nphys265

[16] Urban T, Gassert FT, Frank M, et al. Qualitative and quantitative assessment of emphysema using dark-field chest radiography. Radiology 2022;303:119-127. doi:10.1148/radiol.212025

[17] van Ginneken B, Detio AAA, Jacobs C. Luna16 Dataset. https://luna16.grand-challenge.org/data/. Published 2016. Accessed January 16, 2020.

[18] Colak E, Kitamura FC, Hobbs SB, et al. The RSNA Pulmonary Embolism CT Dataset. Radiol Artif Intell. 2021;3:e200254. doi:10.1148/ryai.2021200254

[19] Virtanen P, Gommers R, Oliphant TE, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods 2020;17:261-272. doi:10.1038/s41592-019-0686-2

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

[21] Bujila R, Omar A, Poludniowski G. A validation of SpekPy: A software toolkit for modelling X-ray tube spectra. Phys Med 2020. doi:10.1016/j.ejmp.2020.04.026

[22] Omar A, Andreo P, Poludniowski G. A model for the energy and angular distribution of x rays emitted from an x-ray tube. Part I. Bremsstrahlung production. Med Phys 2020;47:4763-4774. doi:10.1002/mp.14359

[23] Omar A, Andreo P, Poludniowski G. A model for the energy and angular distribution of x rays emitted from an x-ray tube. Part II. Validation of x-ray spectra from 20 to 300 kV. Med Phys 2020;47:4005-4019. doi:10.1002/mp.14360
[24] Buzug T. From photon statistics to modern cone-beam CT. In: Computed Tomography. Heidelberg, Springer Berlin, 2008, 477.

[25] Mechlem K, Ehn S, Sellerer T, et al. Joint Statistical Iterative Material Image Reconstruction for Spectral Computed Tomography Using a Semi-Empirical Forward Model. IEEE Trans Med Imaging 2018;37:68-80. doi:10.1109/TMI.2017.2726687

[26] Hubbell JH, Seltzer SM. X-ray mass attenuation coefficients. https://www.nist.gov/pml/x-ray-mass-attenuation-coefficients. Published 2004. Accessed January 16, 2020.

[27] Schoonjans T, Brunetti A, Golosio B, et al. The xraylib library for X-ray–matter interactions. Recent developments. Spectrochimica Acta Part B: Atomic Spectroscopy 2011;66: 776-784. doi:10.1016/j.sab.2011.09.011.

[28] van Aarle W, Palenstijn WJ, Cant J, Janssens E, Bleichrodt F, Dabravolski A, De Beenhouwer J, Joost Batenburg K, Sijbers J. Fast and flexible X-ray tomography using the ASTRA toolbox. Opt Express 2016 Oct 31;24:25129-25147. doi: 10.1364/OE.24.025129.

[29] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. 18th International Conference on Medical Image Computing and Computer-Assisted Intervention MICCAI; 2015 Oct 5- Oct 9; Munich, Germany. Springer, Cham; 2015 Nov. Proceedings, Part III 18, 234-241 p. doi: 10.48550/arXiv.1505.04597.

[30] keras.io. Keras Simple. Flexible. Powerful; 2015. Available from: https://keras.io. Accessed Feb 10, 2023.

[31] tensorflow.org. Create production-grade machine learning methods with TensorFlow; 2015. Available from: https://www.tensorflow.org/. Accessed Feb 10, 2023.

[32] Harris, C.R., Millman, K.J., van der Walt, S.J. et al. Array programming with NumPy. Nature 585, 357–362 (2020). DOI: 10.1038/s41586-020-2649-2

[33] Haas M, Hamm B, Niehues SM. Automated lung volumetry from routine thoracic CT scans: how reliable is the result? Acad Radiol 2014;21:633-8. doi: 10.1016/j.acra.2014.01.002.
[34] Wisselink HJ, Steerenberg DJD, Rook M, Pelgrim GJ, Heuvelmans MA, van den Berge M, de Bock GH, Vliegenthart R. Predicted versus CT-derived total lung volume in a general population: The ImaLife study. PLoS One 2023;18:e0287383. doi: 10.1371/journal.pone.0287383.

[35] Danielsbacka JS, Olsén MF, Hansson PO, Mannerkorpi K. Lung function, functional capacity, and respiratory symptoms at discharge from hospital in patients with acute pulmonary embolism: A cross-sectional study. Physiother Theory Pract 2018;34:194-201. doi: 10.1080/09593985.2017.1377331.
Table 1

**Subject Demographics for In-House Dataset (n=72)**

| Parameter \ Subset | Healthy | COPD |
|--------------------|---------|------|
| Male               | 20      | 25   |
| Female             | 13      | 14   |
| Age (years)        | 62.4 [34, 80] | 69.0 [47, 91] |

Note. Age is given in terms of each set’s mean and age range [youngest, oldest].

COPD = Chronic Obstructive Pulmonary Disease
| Dataset \ Metric          | Luna16 (n = 131) | PE (n = 1,060) | Healthy (n = 33) | COPD (n = 39) |
|--------------------------|------------------|----------------|-----------------|--------------|
| Total Lung Volume (L)     | 5.45             | 4.10           | 5.12            | 5.79         |
| [95% CI] (L)              | [5.00 – 5.90]    | [3.96 – 4.24]  | [4.51 – 5.73]   | [5.13 – 6.44]|
| Mean Predicted (L)        | 5.39             | 4.06           | 5.13            | 5.64         |
| [95% CI] (L)              | [4.60 – 5.83]    | [3.93 – 4.20]  | [4.55 – 5.72]   | [5.04 – 6.24]|
| MSE (L)                   | 0.094            | 0.166          | 0.112           | 0.282        |
| MAE (L)                   | 0.230            | 0.243          | 0.250           | 0.379        |
| Pearson Coefficient       | 0.993            | 0.985          | 0.981           | 0.968        |
| (P < 0.001)               |                  |                | (P < 0.001)     | (P < 0.001)  |

CI = Confidence Interval

COPD = Chronic Obstructive Pulmonary Disease

MAE = Mean Absolute Error

MSE = Mean Squared Error

PE = Pulmonary Embolism
### Table 3
**Total Lung Volume for Real Radiographs**

| Dataset \ Metric     | Healthy (n = 33) | COPD (n = 39) | All (n = 72) |
|----------------------|-----------------|---------------|--------------|
| Mean Ground Truth    | 4.85            | 5.51          | 5.21         |
| [95% CI] (L)         | [4.43 – 5.26]   | [5.05 – 5.97] | [4.89 – 5.52]|
| Mean Predicted       | 4.83            | 5.28          | 5.08         |
| [95% CI] (L)         | [4.47 – 5.19]   | [4.93 – 5.64] | [4.82 – 5.33]|
| MSE (L)              | 0.300           | 0.394         | 0.351        |
| MAE (L)              | 0.425           | 0.494         | 0.462        |
| Pearson Coefficient  | 0.881 (P < 0.001)| 0.921 (P < 0.001) | 0.908 (P < 0.001) |

CI = Confidence Interval

COPD = Chronic Obstructive Pulmonary Disease

MAE = Mean Absolute Error

MSE = Mean Squared Error
**Figure 1.** Data selection flowchart. **A)** Public chest CT data from the Luna16 and the RSNA pulmonary embolism (PE) challenges were considered to generate synthetic radiographs. Only scans acquired with 120kVp with a determined diagnosis were selected. Luna16 and PE data were split separately on a 60:20:20 ratio for the train, validation, and test sets. Synthetic radiographs were generated by simulating 10 projections for each sample from -10° to 10° at 2° steps. All projections were utilized for the train and validation sets from the Luna16 data. Only central projections at 0° were chosen for the Luna16 test set, and all of the PE data sets. **B)** Inhouse chest CT scans were retrospectively selected from our clinics, such that all scans were acquired at 120kVp and had a corresponding radiograph scanned no longer than a week apart. Subjects with indeterminate diagnosis CT scans of subjects retrospectively selected from 10.2018 to 12.2019 Subjects with signs of pneumonia, abscess, metastasis, pleural effusion, lung carcinoma, pneumothorax, or other lung diseases Subjects scanned without iodinated contrast material Subjects with indeterminate diagnosis Subjects without a corresponding radiograph or with highly absorbing foreign objects present in their radiographs
Figure 2. Illustrated workflow for generating synthetic radiographs and lung thickness maps for training a U-Net. 

A) The CT scan is segmented into soft tissue, adipose tissue, and bone. An X-ray spectrum is simulated for the forward projection of the CT to generate the synthetic radiograph. To create the two-dimensional ground truth lung thickness density map, the three-dimensional lung segmentation mask is forward projected. The U-Net model is trained with synthetic radiograph input data and corresponding thickness map ground truth labels. 

B) The U-Net model architecture for 256-by-256 pixel input data. The number of feature channels is given above the corresponding block. A three-by-three convolution links a feature block to the next, if not specified otherwise. Upsampling was implemented with nearest neighbor interpolative resizing.
Figure 3. Qualitative results of eight cases from the synthetic radiographs of the public test set, where each column represents one case. A) Simulated input radiographs, B) predicted thickness maps, C) ground truth thickness maps, and D) absolute difference between ground truth and predicted maps. The color bars for (B) and (C) indicate lung thickness in millimeters [mm]. The color bar for (D) indicates the pixel-level estimation error in mm. Cases one through four are from the Luna16 test set, and cases five through eight are from the pulmonary embolism test set.
Figure 4. Qualitative results of eight cases from the in-house test set, where each column represents one case. A) Simulated input radiographs, B) predicted thickness maps for (A), C) ground truth thickness maps for (A), D) absolute difference between ground truth (C) and predicted (B) thickness maps, E) real input radiographs, and F) predicted thickness maps for (E). The CT scan and the lung segmentation were forward projected respectively to obtain (A) and (C). The predicted thickness map for real radiographs (F) and the predicted thickness map for their synthetic pairs (B) can be compared for each case. The color bars for (B), (C), and (F) indicate lung thickness in millimeters [mm]. The color bar for (D) indicates the pixel-level estimation error in mm. Cases one through four are from the healthy set, and cases five through eight are from the chronic obstructive pulmonary disease set.
**Figure 5.** Scatter plots for the ground truth versus predicted total lung volumes in liters [L] for **A)** synthetic radiographs from the public test sets of the Luna16 (orange) and the RSNA pulmonary embolism (PE)(blue) challenges, **B)** synthetic radiographs from our in-house dataset, including healthy subjects (blue), and subjects suffering from chronic obstructive pulmonary disease (COPD)(orange), and **C)** real radiograph pairs for the same subjects in (B). The total number of radiographs is given by \( n \). Pearson’s correlation coefficient, \( r \), and statistical significance are reported. The linear fit is depicted with the black line.

**Figure 6.** Qualitative results of four cases with failed thickness map predictions for real radiographs, where each column represents one case. **A)** input radiographs and **B)** corresponding predicted thickness maps. In all cases, the changes in the intensity range caused by strongly absorbing features in the input radiograph resulted in failed predictions. The color bar for (B) indicates lung thickness in millimeters [mm]. Cases one and two correspond to healthy subjects. Cases three and four correspond to subjects suffering from chronic obstructive pulmonary disease.
Funding:

Funded by the Federal Ministry of Education and Research (BMBF) and the Free State of Bavaria under the Excellence Strategy of the Federal Government and the States, the German Research Foundation (GRK2274), as well as by the Technical University of Munich–Institute for Advanced Study.