TITLE:

A homological approach to a mathematical definition of pulmonary fibrosis and emphysema on computed tomography

AUTHOR(S):

Tanabe, Naoya; Kaji, Shizuo; Sato, Susumu; Yokoyama, Tomoo; Oguma, Tsuyoshi; Tanizawa, Kiminobu; Handa, Tomohiro; Sakajo, Takashi; Hirai, Toyohiro

京都大学
KYOTO UNIVERSITY

american physiological society®

JOURNAL OF APPLIED PHYSIOLOGY®

RESEARCH ARTICLE

A homological approach to a mathematical definition of pulmonary fibrosis and emphysema on computed tomography

Naoya Tanabe,[1]* Shizuo Kaji,[2]* Susumu Sato,[1] Tomoo Yokoyama,[3] Tsuyoshi Oguma,[1] Kiminobu Tanizawa,[1] Tomohiro Handa,[1,4] Takashi Sakajo,[5] and Toyohiro Hirai[1]

[1]*Department of Respiratory Medicine, Graduate School of Medicine, Kyoto University, Kyoto, Japan;* [2]*Institute of Mathematics for Industry, Kyushu University, Fukuoka, Japan;* [3]*Department of Mathematics, Kyoto University of Education, Kyoto, Japan;* [4]*Department of Advanced Medicine for Respiratory Failure, Graduate School of Medicine, Kyoto University, Kyoto, Japan; and* [5]*Department of Mathematics, Kyoto University, Kyoto, Japan*

Abstract

Three-dimensional imaging is essential to evaluate local abnormalities and understand structure-function relationships in an organ. However, quantifiable and interpretable methods to localize abnormalities remain unestablished. Visual assessments are prone to bias, machine learning methods depend on training images, and the underlying decision principle is usually difficult to interpret. Here, we developed a homological approach to mathematically define emphysema and fibrosis in the lungs on computed tomography (CT). With the use of persistent homology, the density of homological features, including connected components, tunnels, and voids, was extracted from the volumetric CT scans of lung diseases. A pair of CT values at which each homological feature appeared (birth) and disappeared (death) was computed by sweeping the threshold levels from higher to lower CT values. Consequently, fibrosis and emphysema were defined as voxels with dense voids having a longer lifetime (birth-death difference) and voxels with dense connected components having a lower birth, respectively. In an independent dataset including subjects with idiopathic pulmonary fibrosis (IPF), chronic obstructive pulmonary disease (COPD), and combined pulmonary fibrosis and emphysema (CPFE), the proposed definition enabled accurate segmentation with comparable quality to deep learning in terms of Dice coefficients. Persistent homology-defined fibrosis was closely associated with physiological abnormalities such as impaired diffusion capacity and long-term mortality in subjects with IPF and CPFE, and persistent homology-defined emphysema was associated with impaired diffusion capacity in subjects with COPD. The present persistent homology-based evaluation of structural abnormalities could help explore the clinical and physiological impacts of structural changes and morphological mechanisms of disease progression.

NEW & NOTEWORTHY This study proposes a homological approach to mathematically define a three-dimensional texture feature of emphysema and fibrosis on chest computed tomography using persistent homology. The proposed definition enabled accurate segmentation with comparable quality to deep learning while offering higher interpretability than deep learning-based methods.

computed tomography; deep learning; fibrosis; lung disease; persistent homology

INTRODUCTION

The lungs are the primary organs of the respiratory system, comprising airways, parenchyma, and vessels, and are continuously involved in gas exchange (1). These structural components are three-dimensionally well organized in healthy lungs, but respiratory diseases disrupt this system, which may cause lung function deterioration and even a poor prognosis (2). Idiopathic pulmonary fibrosis (IPF) is a major fibroproliferative lung disease that induces parenchymal fibrosis (3, 4). Chronic obstructive pulmonary disease (COPD) is also a major lung disease characterized by parenchymal destruction (emphysema) and airway disease (5).

These chronic lung diseases induce heterogeneous structural alterations closely associated with local dynamic changes in molecular pathways and underlie the variability in clinical and physiological outcomes. Therefore, a detailed quantitative morphological evaluation is critical to reveal the pathogenesis of these diseases, understand the structure-function relationships in the lungs, and improve outcomes.

Medical three-dimensional imaging is essential for diagnosing diseases and evaluating local structural abnormalities. Computed tomography (CT) images comprise voxels that reflect tissue density measured in CT values and are widely used to evaluate lung diseases (6, 7). Visual assessments by expert radiologists are often regarded as the gold

standard for morphological diagnosis, and quantitative methods have been explored to aid these visual assessments. Emphysema is characterized by lower CT values due to parenchymal destruction (6), whereas fibrosis is characterized by higher CT values due to deposition of extracellular matrix such as collagen (7, 8). However, this notion is an oversimplification; thus, the quality of simple thresholding segmentation is unsatisfactory. Simple thresholding cannot accurately segment honeycomb cysts, which are a typical radiological finding of IPF characterized by low-density cystic regions surrounded by walls with higher density. Additionally, fibrosis and emphysema coexist in subjects diagnosed with combined pulmonary fibrosis and emphysema (CPFE) (9, 10). Texture analysis and deep learning (DL) techniques have been investigated to overcome these problems (10–12). However, they require many consistently labeled images that radiologists manually and laboriously prepare (13), and the underlying decision principle for these data-driven "black box" methods is usually difficult to interpret. Therefore, a reproducible and interpretable mathematical definition of local radiological abnormalities remains an unmet need to facilitate the use of three-dimensional imaging and improve clinical management.

Honeycomb cysts, for example, have a characteristic three-dimensional structure, as the name suggests. Thus various radiological findings can be characterized by topological features. A three-dimensional shape can be extracted from a volumetric CT by thresholding at a specific CT value. The topological properties of the three-dimensional shape are captured by a tool in topological data analysis called homology, which describes the number of holes of a given dimension such as the connected components (dimension 0), tunnels (dimension 1), and voids (dimension 2). However, different thresholds yield different outputs. This difficulty was overcome by extending the concept of homology to persistent homology, which describes the persistence of homological features (14). Persistent homology encodes a pair of the level at which a homological feature (cycle) of each dimension appears (birth) and the level at which it disappears (death) by sweeping the threshold level from high to low values. The obtained persistent homology data can be visualized using a persistence diagram in which each homological cycle is represented by a point with coordinates (death, birth) on the plane according to the dimensions (15, 16). In medical imaging, persistent homology has allowed the detection of a difference in the network structure of airway trees on CT between patients with COPD and non-COPD controls (17) and the classification of abnormal liver lesions on CT and MRI that conventional methods could not detect (18, 19). However, the previously reported methods exclusively focused on the overall information of a given image (the whole volume or a small patch) but have not yet produced localized information on radiological abnormalities.

The goal of this collaborative work by experts in medicine and mathematics is to establish a homological approach to a mathematical definition of any specific local regions on volumetric imaging beyond the global signatures, allowing segmentation of the complicated structures and uncovering the mechanism that governs the local structural alteration and functional abnormalities. As a first step, the present study specifically focuses on defining pulmonary fibrosis and emphysema on CT using persistent homology to localize these abnormalities with a comparable accuracy to DL-based segmentation while providing simple and interpretable definitions that are difficult to generate using DL-based methods. Persistent homology-based definitions of fibrosis and emphysema were established using the first dataset of CT scans from controls and COPD and IPF subjects and then were validated in the independent second dataset of CT scans from COPD, IPF, and CPFE subjects.

METHODS

Study Design

Two different datasets were prepared. Among the subjects who had undergone inspiratory CT scans at full inspiration and spirometry at the hospital from 2006 to 2014, we randomly selected 15 healthy nonsmokers (controls), 15 subjects with COPD, and 15 subjects with IPF without visual CT findings of emphysema for the first dataset and 30 COPD subjects, 30 IPF subjects without visual CT findings of emphysema, and 30 CPFE subjects for the second dataset. All the subjects with COPD and CPFE were smokers. Control subjects were defined as nonsmokers aged 30 years or older with normal spirometry, no abnormal CT findings on visual inspection, and no history of lung disease. The exclusion criteria for subjects with COPD, CPFE, and IPF were 1) age younger than 40 yr, 2) a history of lung resection surgery and radiation therapy to the lungs, and 3) other lung diseases, such as lung cancer, bronchiectasis, and chronic lung infection. The diagnosis of IPF was based on multidisciplinary discussion (MDD) according to the Official ATS/ERS/JRS/ALAT Clinical Practice Guideline for the diagnosis of IPF (20). COPD was diagnosed based on a combination of post-bronchodilator forced expiratory volume in 1 s (FEV_1)/forced vital capacity (FVC) <0.7, smoking history >10 pack-yr, and respiratory symptoms (5). All COPD subjects showed visual CT findings of emphysema without fibrotic changes. CPFE was defined as subjects with IPF who showed visual CT findings of emphysema. Spirometry and measurement of the carbon monoxide diffusing capacity (D_{LCO}) were performed using the single-breath method and the Chestac-65V system (Chest MI Corp.). The composite physiological index (CPI), which has been used as a major index to estimate fibrosis, was calculated using the following equation: $91.0 - [0.65 \times \%predicted\ D_{LCO}] - [0.53 \times \%\ of\ predicted\ FVC] + [0.34 \times \%\ of\ predicted\ FEV_1]$ (21). The prognostic information after CT scans over 5 yr was retrospectively reviewed in the medical records. The Ethics Committee approved the retrospective analysis of the data (no. R1323) and waived the written informed consent requirement.

CT Acquisitions and Lung Segmentations

CT scans were performed with Aquilion One and Prime scanners (Canon Medical Systems) at the 120-kV peak (kVp) and autoexposure control, and images with a 0.5-mm slice thickness were reconstructed using a sharp reconstruction algorithm (FC51 or 56) for all the participants in the two datasets. The fields of view were 350 mm and 320 mm, and the voxel resolutions were $0.683 \times 0.683 \times 0.5$ mm and $0.625 \times 0.625 \times 0.5$ mm for the male and female subjects,

京都大学
KYOTO UNIVERSITY

respectively. From the original CT images, the lungs were segmented using deep neural networks according to a report by Kaji et al. (22) (https://github.com/shizuo-kaji/PairedImageTranslation), who provided a Python code for generic image translation algorithms, including widely used U-Net (23) with several enhancements, such as GAN regularizers. With this script, a segmentation model was trained by a medical doctor using 988 pairs of original CT images and segmented lung images that were obtained from an independent dataset of participants for the present analyses. The lung fields were then extracted from the CT images based on the trained model. The same script was used for the DL-based segmentation of fibrosis and emphysema as a comparison target to the proposed method. A neural network model based on U-Net (23) was trained using a set of 114 manually segmented axial slices chosen from the first dataset. The validation of the present DL method was performed using the manual segmentation of CT images (25 slices) randomly selected from the second dataset as the reference in this study. These selected images were also used to evaluate the performance of persistent homology-based segmentation. Additionally, the conventional threshholding method defined fibrosis as CT voxels with greater than -200 HU (24) and emphysema as CT voxels with less than -50 HU (25).

Persistent Homology Computation to Obtain a Persistence Diagram to Visualize a Homological Feature of Volumetric CT Data

Homological features of a given volumetric CT data were extracted by persistent homology using the open source program Cubical Ripser (https://github.com/shizuo-kaji/CubicalRipser_3dim) (26). The mathematical details are

described in the APPENDIX. Each three-dimensional volume of the segmented lung region provided a triple persistence diagram, in which the CT values for the birth and death of connected components (H_0), tunnels (H_1), and voids (H_2) were visualized separately (Fig. 1A). Each point in the plot at (x, y) in the persistence diagram represents a topological feature that emerges at the CT value y and disappears at the CT value x.

Unsupervised Analysis of the Persistence Diagram Using Principal Component Analysis

To confirm the potential of the persistence diagrams representing homological features of volumetric CT to differentiate different lung diseases, triple persistence diagrams (H_0, H_1, and H_2) were vectorized (600 dimensions) and processed with principal component analysis (PCA).

Persistence Diagrams of the Cubic Regions of Interest for Fibrosis, Emphysema, and Normal Regions on CT

To explore the rigorous definitions of fibrosis and emphysema regions on CT using persistent homology, cubic regions of interest (ROIs; $25 \times 25 \times 25$ voxels; $n = 543$) for fibrosis ($n = 132$), emphysema ($n = 182$), and normal regions ($n = 229$) were selected from the volumetric CT of the 45 subjects in the first dataset and computed to generate the persistence diagram for each ROI (Fig. 1B). Based on the guidelines for the diagnosis of IPF (20), regions with honeycomb cysts and reticular shadows, but not pure ground-glass opacification (GGO), were identified as ROIs for IPF-related fibrosis. Because the voxel resolutions were $0.683 \times 0.683 \times 0.5$ mm and $0.625 \times 0.625 \times 0.5$ mm for the male and female subjects, respectively, the sizes of the ROIs were $17.1 \times 17.1 \times 12.5$ mm and $15.6 \times 15.6 \times 12.5$ mm for the male and female
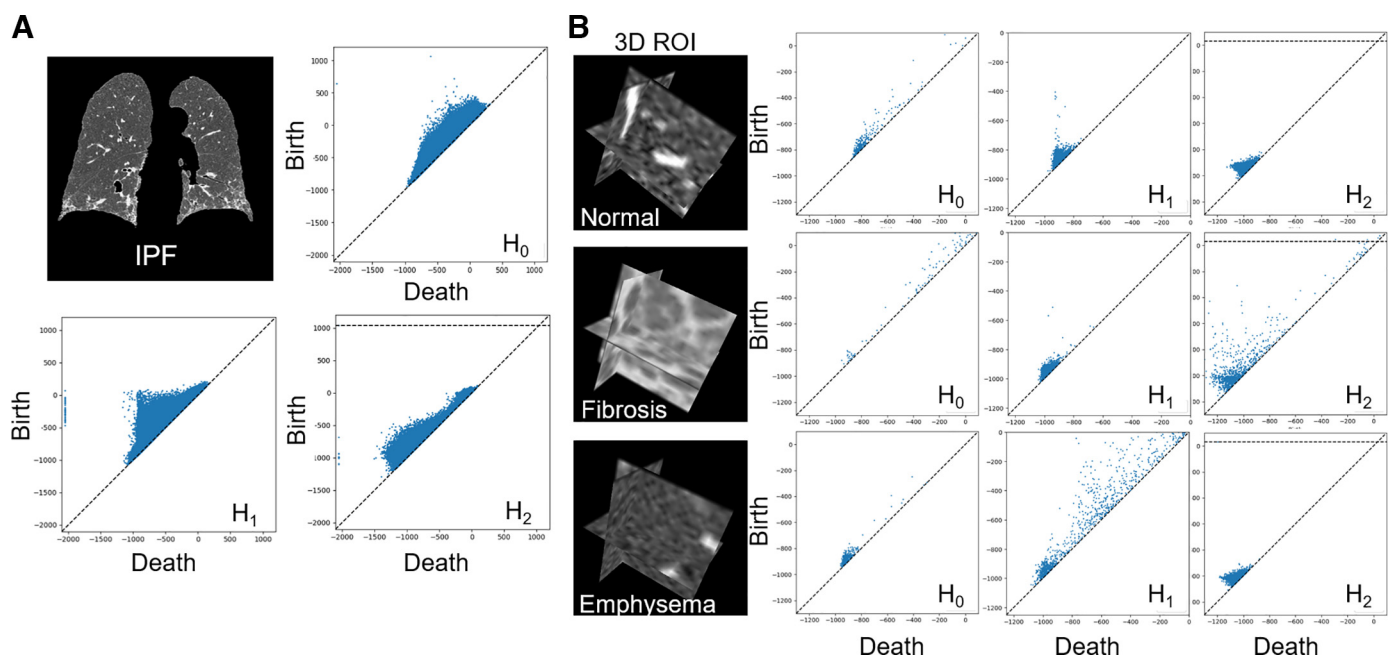

Figure 1. Computed tomography images of whole lungs and regions of interest and persistence diagram. *A*: an example of a coronal computed tomography image with idiopathic pulmonary fibrosis (IPF) and persistence diagrams including dimensions 0 (H_0), 1 (H_1), and 2 (H_2). *B*: normal, fibrosis, and emphysema regions of interest (ROIs) were selected from the first dataset of control, IPF, and chronic obstructive pulmonary disease lungs, respectively. The size of the ROI was $25 \times 25 \times 25$ voxels. The persistence diagrams were computed for dimensions 0 (H_0), 1 (H_1), and 2 (H_2).

A Self-archived copy in
Kyoto University Research Information Repository
https://repository.kulib.kyoto-u.ac.jp

京都大学
KYOTO UNIVERSITY

KURENAI 紅
Kyoto University Research Information Repository

HOMOLOGY AND RADIOLOGICAL ABNORMALITIES IN LUNG DISEASES

subjects, respectively. Generally, the size of individual honeycomb cysts is 2–3 mm in diameter, although larger honeycomb cysts may be present in some IPF lungs (7, 27). Therefore, the $25 \times 25 \times 25$ voxel ROI size was chosen because the corresponding sizes ($17.1 \times 17.1 \times 12.5$ mm and $15.6 \times 15.6 \times 12.5$ mm) were considered to include multiple honeycomb cysts and allow robust segmentations for cysts.

Decision Tree Models to Classify ROIs into Fibrosis, Emphysema, and Normal Regions by Their Persistence Diagrams

Because it is clinically important to find a small set of interpretable variables that considerably affect the diagnosis, a single set of variables on the persistence diagram for each fibrosis and emphysema that can discern the type of lesion was searched. For this purpose, decision tree models were trained to classify ROIs (normal, fibrosis, and emphysema) based on persistence diagrams. After visually labeling all the ROIs, an automatic grid search was performed to identify the dimensions (H_0, H_1, or H_2), birth, and lifetime (difference between birth and death) characterizing cycles for fibrosis and emphysema on the persistence diagram with the highest classification accuracy.

Calculation of the Cycle Density for Fibrosis and Emphysema for Each CT Voxel

By using the definition of the cycles for fibrosis and emphysema on persistence diagrams, the density of these cycles in the neighborhood of each voxel was measured throughout the lungs. More specifically, for each voxel, the sum of the number of fibrotic (emphysematous) cycles weighted by the Gaussian kernel $e^{\frac{-u^2}{12}}$ was computed, where u denotes the distance between the voxel and birth location of the cycle. In this equation, we chose 12 as the window parameter in conjunction with the general size of honeycomb cysts (2–3 mm in diameter) (7, 27). Following the calculation of the cycle density, each voxel was assigned to fibrosis, emphysema, and normal voxels based on the threshold values for the cycle density for fibrosis and emphysema.

Dice Coefficients

The segmentation quality was evaluated by the Dice coefficients (28), which quantified the similarity between a given segmentation method and the manual segmentation that was considered the reference. Interobserver variability was assessed by comparing the reference manual segmentation and the other manual segmentation by an independent analyst.

Statistics

Statistical analyses were performed in R (29). Multiple comparisons of the cycle density, persistent homology-defined fibrosis and emphysema were performed with one-way ANOVA followed by Tukey's test and the Kruskal-Wallis test followed by Dunn's test (30). The correlations were tested using the Pearson and Spearman correlation tests, and the mortality between two groups was evaluated using Kaplan-Meier survival curves with log-rank tests.

RESULTS

Table 1 shows the demographics of the two independent datasets, including control, IPF without emphysema, and COPD (n = 15 for each) in the first dataset and IPF without emphysema, CPFE, and COPD (n = 30 for each) in the second dataset. The severity of COPD assessed by FEV_1 (%predicted) did not differ between the first and second datasets (83% and 85%; P = 0.20), and the severity of IPF assessed by D_{LCO} (% predicted) and CPI did not differ between them (D_{LCO}: 50% and 46%, P = 0.47; CPI: 47 and 48, P = 0.85).

The persistence diagrams for H_0, H_1, and H_2 were obtained from all the CT scans in the first dataset, vectorized, and processed with PCA (Fig. 1). The explained variances of the first two components were 73.58% and 23.63%, with a total of 97.17%, and each case was plotted as a point using the two components (Fig. 2A). The plot showed that points from the IPF and COPD subjects and controls were separated from each other. Furthermore, each case from the second dataset was plotted using the first and second principal components

Table 1. *Demographics of the two datasets*

	1st Data (Discovery)			**2nd Dataset (Validation)**		
	Control	**IPF**	**COPD**	**CPFE**	**IPF**	**COPD**
n	15	15	15	30	30	30
Age, yr	62±12	64±10†	74±6	74±8	73±8†	70±10
Male, %	27%	67%	100%	80%	90%	90%
Height, cm	156±10	163±7	165±6	162±7	164±6	161±6
Weight, kg	56±11	64±13	57±6	62±12	63±12	59±10
Smoking, pack-yr	0±0	45±52	53±23	59±26	29±20	61±30
FVC (%predicted), %	97±9	73±16	93±16	93±20	82±22	90±18
FEV_1 (%predicted), %	96±10	83±14	54±18	85±19	90±22	58±20
FEV_1/FVC, %	78±4	89±6	45±12	72±14	85±7	50±15
D_{LCO}, %predicted*	NA	50±13	53±21	42±12	46±16	58±16
CPI*	NA	47±12	26±13	43±12	48±14	25±11

Values are means ± SD. FEV_1, forced expiratory volume in 1 s; FVC, forced vital capacity. *Diffusion capacity of the lung for carbon monoxide (D_{LCO}) was measured in 14 idiopathic pulmonary fibrosis (IPF) and 15 combined pulmonary fibrosis and emphysema (COPD) subjects but not in the controls in the first dataset and in 29 IPF, 29 combined pulmonary fibrosis and emphysema (CPFE), and 30 COPD subjects in the second dataset. The composite physiological index (CPI) was calculated in subjects whose D_{LCO} was measured. †Age in subjects with IPF significantly differed between the 2 groups (P = 0.001), while other variables did not differ between the datasets in subjects with IPF and COPD.
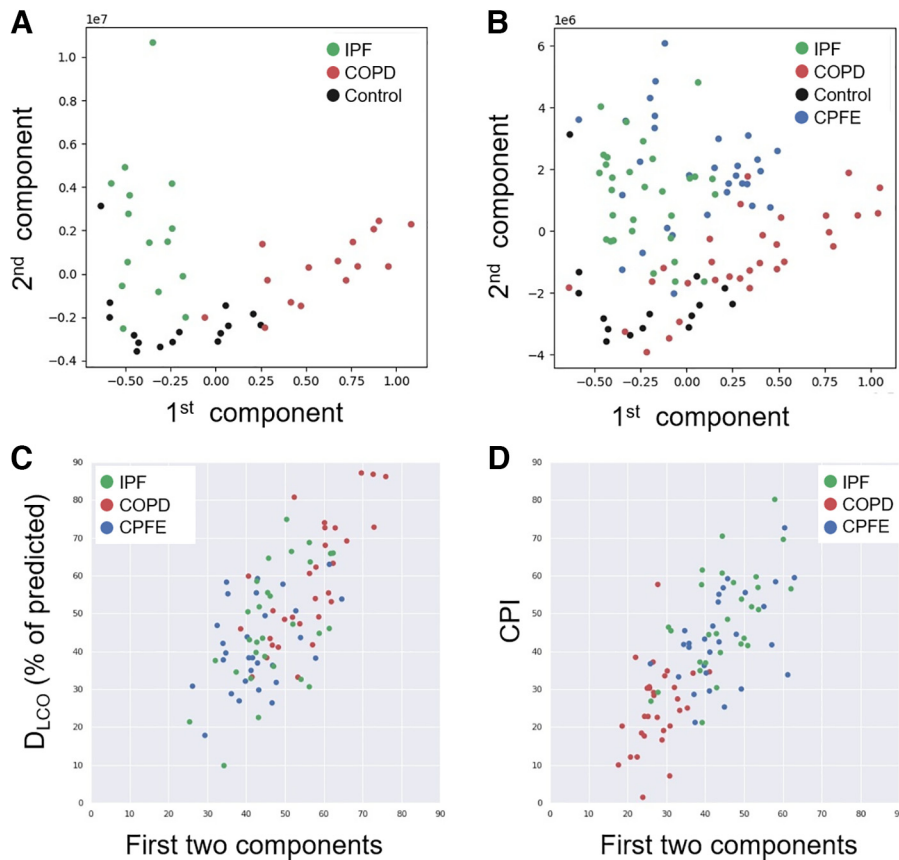
Figure 2. Principal component analysis of vectorized persistence diagrams from different diseases was performed. Each case in the 1st dataset was plotted using the 1st and 2nd principal components (*A*). In the 2nd dataset, with the use of the 1st and 2nd principal components obtained in the 1st dataset, each case from the 2nd dataset was plotted (*B*). Associations of the 1st 2 principal components in the second dataset with diffusion capacity for carbon monoxide (D_{LCO}) and the composite physiological index (CPI) are shown (*C* and *D*). With the use of the coordinates of the 2 principal components as explanatory variables, a linear regression model for the target variable was fitted for the second dataset. The *x*-axis indicates the predicted value, and the *y*-axis indicates the real target value. IPF, idiopathic pulmonary fibrosis; COPD, chronic obstructive pulmonary disease; CPFE, combined pulmonary fibrosis and emphysema.

obtained in the first dataset (Fig. 2*B*). There was a tendency toward separation between the plots from the CPFE, COPD, and IPF subjects, although the first dataset did not contain CPFE subjects. Additionally, the first two principal components were associated with D_{LCO} ($r = 0.67$, $P < 0.001$) and CPI ($r = 0.72$, $P < 0.001$) (Fig. 2, *C* and *D*). These data suggest that the characteristics of CT images for each disease are encoded in the persistence diagram.

The decision tree models classified the persistence diagrams of ROIs for fibrosis ($n = 132$), emphysema ($n = 182$), and normal regions ($n = 229$) (Fig. 3*A*). The ROI for fibrosis was accounted for by cycles with dimension = 2, $-1,260 < \text{birth} < -380$, and $360 < \text{lifetime} < \infty$. The ROI for emphysema was accounted for by cycles with dimension = 0, $-1,020 < \text{birth} < -900$, and $20 < \text{lifetime} < 90$. Furthermore, when the cycle density for fibrosis in the neighborhood of a given voxel was above the threshold of 1.0, the voxel was considered persistent homology-defined fibrosis. For the remaining voxels, when the cycle density for emphysema in the neighborhood of a given voxel was above the threshold of 8.3, the voxel was considered persistent homology-defined emphysema. The confusion matrix in Fig. 3*A* shows that no ROI for fibrosis was assigned to persistent homology-defined emphysema and no ROI for emphysema was assigned to persistent homology-defined fibrosis. The sensitivities of persistent homology-based classification of fibrosis and emphysema were 0.93 and 0.89, and the specificities were 0.97 and 0.85, respectively. Figure 3*B* compares the persistence diagrams with the established definition of the cycles for fibrosis and emphysema in

normal, fibrosis, and emphysema ROIs. Figure 3*C* shows examples of the localization of persistent homology-defined fibrosis and emphysema on CT. The volume percentage of persistent homology-defined fibrosis to the whole lung (PH-fibrosis%) was significantly higher in the subjects with IPF than in the controls and subjects with COPD ($16.60 \pm 0.130\%$, $0.69 \pm 0.01\%$, and $0.40 \pm 0.00\%$, respectively), whereas the volume percentage of persistent homology-defined emphysema (PH-emphysema%) was significantly higher in the subjects with COPD than in the controls and subjects with IPF ($17.60 \pm 0.20\%$, $0.03 \pm 0.00\%$, and $0.13 \pm 0.00\%$, respectively).

Next, the validity of the persistent homology-based localization of fibrosis and emphysema was confirmed in the second dataset that included 30 CPFE subjects, 30 IPF subjects without visual emphysema on CT, and 30 COPD subjects. The persistent homology-based segmentations identified fibrosis and emphysema regions in a CPFE subject, although CPFE subjects were not included in the first dataset (Fig. 4). The segmentation quality for the persistent homology-based method on the CT scans of IPF and CPFE lungs was comparable to that for the DL-based method and superior to that for the conventional thresholding method. Notably, honeycomb regions were clearly differentiated from emphysematous regions by persistent homology-based segmentation and DL-based segmentation, whereas thresholding misclassified the true honeycomb regions as emphysema regions.

Table 2 compares the segmentation quality between different segmentation methods using manual segmentation as the reference. Representative images are shown in Supplemental
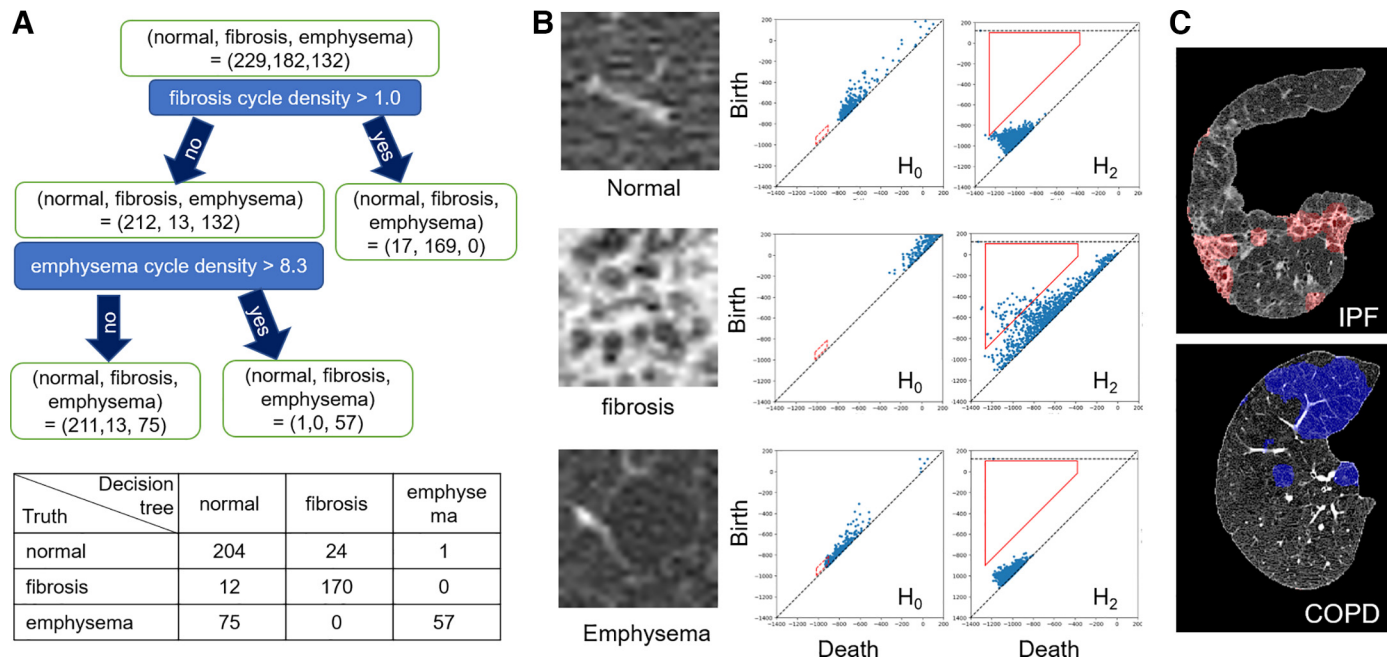
Figure 3. A decision tree model was established to define persistent homology-based fibrosis and emphysema and representative regions of interest (ROIs) based on a persistent diagram. *A*: persistence diagram. The persistent homology-based definition of fibrosis identified 169 of the 182 ROIs with fibrosis. The persistent homology-based definition of emphysema was then applied to identify ROIs with emphysema. *B*: the persistence diagrams of 3 ROIs representing normal, fibrosis and emphysema showed different distributions of cycles on their persistence diagrams. The red solid regions indicate the definition of the fibrosis cycle on the dimension 2 diagram (H_2). The red dashed regions indicate the definition of the emphysema cycle on dimension 0 (H_0). *C*: examples of the persistent segmentation of fibrosis (red) in idiopathic pulmonary fibrosis (IPF) lungs and emphysema (blue) in chronic obstructive pulmonary disease (COPD) lungs.

Fig. S1 (all Supplemental material is available at https://doi.org/10.6084/m9.figshare.14554170). The PH-fibrosis% and DL-based fibrosis% were similarly correlated with fibrosis% measured using the reference manual segmentation (ρ = 0.85, $P < 0.001$ and ρ = 0.87, $P < 0.001$). The PH-emphysema% and DL-based emphysema% were also similarly correlated with emphysema% measured using the reference manual segmentation (ρ = 0.93, $P < 0.001$ and ρ = 0.93, $P < 0.001$). Moreover, the Dice coefficients for segmentation of fibrosis and emphysema in the persistent homology-based method were comparable to those in the DL-based method and the independent analyst's manual segmentation representing interobserver variability.

Table 3 compares fibrosis% and emphysema% measured using different segmentation methods in the second dataset. PH-fibrosis% and DL-based fibrosis% were comparable to
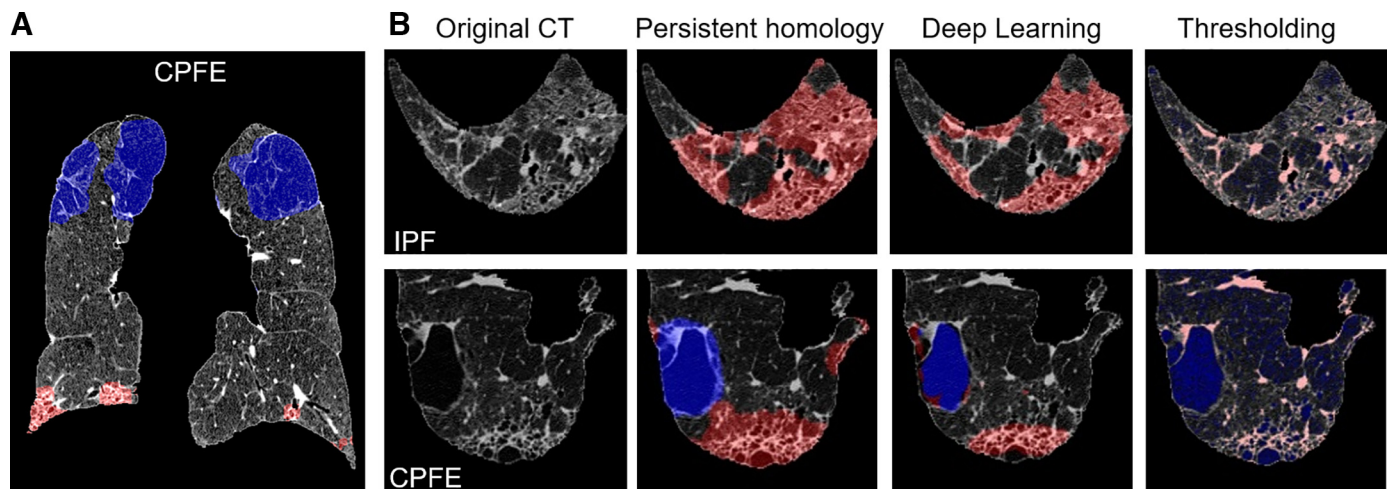
Figure 4. *A*: an example of clear separation of fibrosis (red) and emphysema (blue) regions on computed tomography (CT) from the combined pulmonary fibrosis and emphysema (CPFE) in the second dataset using persistent homology (PH). *B*: persistent homology-based segmentation was comparable to the deep learning-based segmentation method and superior to conventional thresholding. Red and blue colors indicate regions classified as fibrosis and emphysema based on each segmentation method. IPF, idiopathic pulmonary fibrosis.

京都大学
KYOTO UNIVERSITY

KURENAI 紅
京都大学学術情報リポジトリ
Kyoto University Research Information Repository

HOMOLOGY AND RADIOLOGICAL ABNORMALITIES IN LUNG DISEASES

Table 2. *Correlation coefficients for fibrosis and emphysema percentages and Dice coefficients between the manual segmentation and each segmentation method*

	Persistent Homology	Deep Learning	Thresholding	Independent Manual
Correlation coefficients (each method vs. manual)				
Fibrosis%	ρ = 0.86	ρ = 0.85	ρ = 0.84	ρ = 0.86
	$P < 0.001$	$P < 0.001$	$P < 0.001$	$P < 0.001$
Emphysema%	ρ = 0.93	ρ = 0.93	ρ = 0.79	ρ = 0.94
	$P < 0.001$	$P < 0.001$	$P < 0.001$	$P < 0.001$
Dice coefficients (each method vs. manual)				
Fibrosis	0.46 ± 0.14	0.59 ± 0.11	0.11 ± 0.05	0.58 ± 0.14
Emphysema	0.46 ± 0.18	0.38 ± 0.16	0.18 ± 0.08	0.46 ± 0.18

Values are means ± SD. Correlation coefficients were measured using Spearman correlation tests. The percentages of fibrosis and emphysema were compared between each segmentation method and the reference manual segmentation. High and low attenuation regions were segmented as fibrosis and emphysema in the thresholding method. An independent analyst performed manual segmentations of fibrosis and emphysema, which were compared with the reference manual segmentation to estimate the interobserver variability.

each other and significantly higher than the high attenuation area% measured with thresholding in subjects with IPF and CPFE. PH-emphysema% and DL-based emphysema% were comparable to each other and lower than the low attenuation area% measured with thresholding in subjects with COPD.

Table 4 shows that PH-emphysema% was associated with D_{LCO} in subjects with COPD (ρ = −0.66, $P < 0.001$). PH-fibrosis% was associated with D_{LCO} and CPI in subjects with IPF (ρ = −0.71, $P < 0.001$, and ρ = 0.63, $P < 0.001$), whereas PH-fibrosis% was associated with CPI but not with D_{LCO} in subjects with CPFE (ρ = 0.59, $P < 0.001$, and ρ = −0.31, $P = 0.10$). The sum of PH-fibrosis% and PH-emphysema% was associated with D_{LCO} in the subjects with IPF, CPFE, and COPD ($r = −0.58$, $P < 0.001$) (Fig. 5A). The CPI was also associated with PH-fibrosis% in the subjects with IPF and CPFE ($r = 0.70$, $P < 0.001$) (Fig. 5B). The 5-yr mortality rate did not differ between the subjects with CPFE and IPF (Fig. 5C). In contrast, when subjects with CPFE and IPF were divided into a high-persistent homology-defined group and a low-persistent homology-defined group based on PH-fibrosis%, the high-persistent homology-defined fibrosis group showed an increase in mortality compared with the low-persistent homology-defined fibrosis group (Fig. 5D).

Table 3. *Comparisons of the extents of fibrosis and emphysema quantified with different segmentation methods in the second dataset*

	Persistent Homology	Deep Learning	Thresholding
Fibrosis%			
IPF	11.8 ± 9.1*	11.3 ± 5.3*	4.7 ± 2.2
CPFE	9.3 ± 10.1*	9.9 ± 5.7*	3.9 ± 2.2
COPD	0.4 ± 0.5*†	2.3 ± 0.8*	1.7 ± 0.3
Emphysema%			
IPF	0.2 ± 0.4*	0.9 ± 1.2*	14.2 ± 4.6
CPFE	8.5 ± 8.9*	4.4 ± 3.7*	21.1 ± 6.7
COPD	8.9 ± 12.2*	4.2 ± 4.8*	26.0 ± 7.6

Values are means ± SD. Multiple comparisons of fibrosis% and emphysema% among persistent homology, deep learning, and thresholding were performed using Tukey's method. IPF, idiopathic pulmonary fibrosis; CPFE, combined pulmonary fibrosis and emphysema; COPD, chronic obstructive pulmonary disease. *$P < 0.05$, compared with thresholding. †$P < 0.05$, compared with deep learning.

DISCUSSION

The present study used persistent homology to mathematically define two major lung abnormalities, fibrosis and emphysema, on volumetric CT. The analyses of the first dataset established the specific cycles on the persistence diagrams responsible for fibrosis and emphysema, and then the validity of the definition was examined from both a radiological perspective and associations with physiological impairments and clinical outcomes using the independent second dataset. The proposed method is an important step for accurate computerized evaluation of lung fibrosis that can manifest on CT with low attenuation structures (traction bronchiectasis and honeycomb cysts), which is difficult to distinguish from emphysema. Additionally, the efficient segmentations of fibrosis and emphysema in this study suggest that the principles of persistent homology have the potential to characterize local abnormalities and may provide new topological insights into the progression of emphysematous and fibrotic abnormalities.

In the present study, independent datasets were used for development and validation to demonstrate the generalizability of the proposed method. Although automatic segmentation of emphysematous regions within fibrotic regions in CPFE lungs is considered challenging (10), the proposed homological definitions of fibrosis and emphysema using persistent homology have demonstrated notable capacity to differentiate among CPFE, COPD and IPF. Additionally, persistent homology-based segmentations were validated by confirming that PH-emphysema% was associated with D_{LCO} in subjects with COPD, whereas PH-fibrosis% was associated with D_{LCO} and CPI in subjects with IPF. Notably, the finding that PH-fibrosis% was significantly associated with CPI but not with D_{LCO} in subjects with CPFE is in line with a previous paper by Wells et al. (21), who established CPI to estimate the extent of fibrosis on CT in IPF subjects with emphysema more accurately than D_{LCO}.

Furthermore, significant associations were found between D_{LCO} and the sum of PH-fibrosis% and PH-emphysema% in the subjects with CPFE, COPD, and IPF, as well as between the CPI and PH-fibrosis% in the subjects with CPFE and IPF. While the 5-yr mortality rate did not differ between the subjects with CPFE and IPF, persistent homology-defined fibrosis accounted for mortality in these populations. This finding is consistent with a previous report that the visual CT findings of emphysema did not affect mortality in

京都大学
KYOTO UNIVERSITY

A Self-archived copy in
Kyoto University Research Information Repository
https://repository.kulib.kyoto-u.ac.jp

京都大学学術情報リポジトリ
KURENAI 紅
Kyoto University Research Information Repository

HOMOLOGY AND RADIOLOGICAL ABNORMALITIES IN LUNG DISEASES

Table 4. *Physiological impacts of persistent homology-defined fibrosis and emphysema in subjects with COPD, CPFE, and IPF in the second dataset*

	Persistent Homology Fibrosis			Persistent Homology Emphysema		
	CPFE	IPF	COPD	CPFE	IPF	COPD
D_{LCO}	$\rho = -0.31$ $P = 0.10$	$\rho = -0.71$ $P < 0.001$	$\rho = 0.05$ $P = 0.77$	$\rho = -0.05$ $P = 0.79$	$\rho = -0.23$ $P = 0.23$	$\rho = -0.66$ $P < 0.001$
CPI	$\rho = 0.59$ $P < 0.001$	$\rho = 0.63$ $P < 0.001$	$\rho = -0.05$ $P = 0.78$	$\rho = -0.05$ $P = 0.79$	$\rho = 0.25$ $P = 0.20$	$\rho = 0.29$ $P = 0.13$

Values are the Spearman correlation coefficients and *P* values. IPF, idiopathic pulmonary fibrosis; CPFE, combined pulmonary fibrosis and emphysema; COPD, chronic obstructive pulmonary disease; D_{LCO}, diffusion capacity of the lung for carbon monoxide; CPI, composite physiological index.

subjects with IPF (10). Collectively, these findings suggest that the persistent homology-based evaluation of structural abnormalities in diseases should be promising for future physiological investigation and clinical use.

The persistent homology-based analysis methodology developed in this study provides general techniques for volumetric image analysis not restricted to medical sciences. Persistent homology in its original form focuses on the global structure; local information is not captured. To localize the topological feature of the image, we introduced the idea of cycle density utilizing the location of persistent homology cycles. Furthermore, a method utilizing a shallow decision tree was introduced to achieve a simple and interpretable definition of the characteristic cycles. The persistent homology output is a persistence diagram that cannot be directly used with standard statistical techniques. Vectorization (31) is often performed to convert a persistence diagram into a vector as a preprocess for further analysis. Vectorization was used to assess the relevance of persistent homology to the present study by investigating the persistence diagram of the whole lung in the first dataset in an unsupervised manner. However, simplicity and interpretability, which are critical in medical applications, may be lost with vectorization. Thus we

developed a novel analysis using a shallow decision tree. Overall, statistics and domain expertise were combined to provide a definition of characteristic cycles, while the use of domain expertise was restricted to the selection of the ROIs, limiting the effect and source of bias from humans.

Three-dimensional computation of volumetric CT data was performed to directly generate persistence diagrams encoding local topological features. Unsupervised analysis with the first dataset showed that the controls, IPF subjects, and COPD subjects were discernable by the persistence diagrams of the whole lung CT images, suggesting that the entire collection of births and deaths of connected components (H_0), tunnels (H_1), and voids (H_2) could carry distinct information on the radiological features for each disease. This finding is in sharp contrast to the fact that DL-based methods depend on numerous "training images" generated through two-dimensional visual inspection by expert radiologists.

The fibrotic regions were defined as voxels whose neighboring voxels contained cycles with longer lifetimes on the persistence diagrams of dimension 2 (H_2). To interpret the persistent homology-defined fibrosis radiologically, the representative CT ROIs were binarized using different thresholds (Fig. 6). The H_2 cycle (void) was identified on binary images of

Figure 5. Physiological and clinical impacts of the volume percentages of persistent homology-based fibrosis and emphysema (PH-fibrosis% and PH-emphysema%) were evaluated in the 2nd dataset. *A*: the sum of PH-fibrosis% and PH-emphysema% was associated with the diffusion capacity of the lung for carbon monoxide (D_{LCO}) in subjects with idiopathic pulmonary fibrosis (IPF), combined pulmonary fibrosis and emphysema (CPFE), and chronic obstructive pulmonary disease (COPD) (*n* = 30 for each group). *B*: the composite physiological index (CPI) was associated with PH-fibrosis% in subjects with IPF and CPFE. *C*: the 5-yr mortality rate did not differ between subjects with CPFE and IPF. *D*: however, when subjects with CPFE and IPF were divided based on the median PH-fibrosis%, a high-PH-fibrosis group showed an increase in mortality compared with a low-PH-fibrosis group.
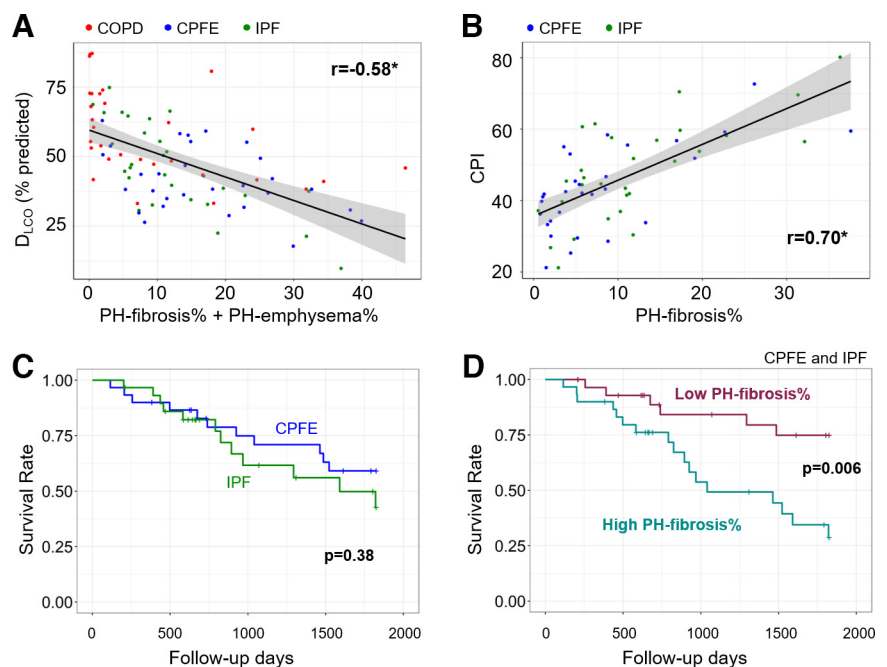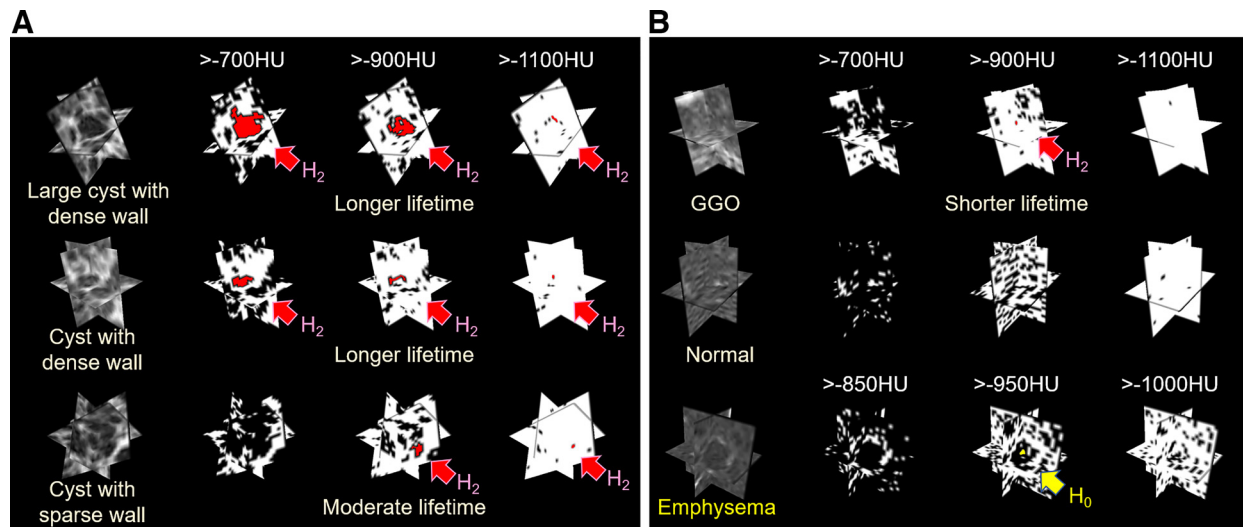
Figure 6. Persistence of voids and connected components on binarized images of honeycomb cysts, ground-glass opacification, emphysema, and normal regions using different computed tomography value thresholds. Cubic regions of interest (ROIs) for fibrotic regions, including a large honeycomb cyst with a dense wall, a cyst with a dense wall, and a cyst with a sparse wall (A), as well as ground-glass opacification (GGO), a normal control region, and emphysema (B), were extracted from volumetric computed tomography (CT). Each CT ROI was binarized with a given threshold of CT value, such as greater than −700 HU, greater than −850 HU, greater than −900 HU, greater than −1,000 HU, and greater than −1,100 HU. Notably, the void components (H_2) persisted at all thresholds of greater than −700 HU, greater than −900 HU, and greater than −1,100 HU for cysts with dense walls regardless of the size of the cysts but not for cysts with sparse walls or GGOs. In contrast, the connected component (H_0) appeared at a lower CT value (in this case, −950 HU) in emphysema.

honeycomb cysts with both dense and sparse walls and GGOs when the threshold CT value was set to greater than −900 HU. In contrast, when the threshold was changed to greater than −1,000 HU, the H2 cycle was identified on binary images of cysts with both dense and sparse walls but not GGOs. These changes resulted in longer and moderate lifetimes in cysts with dense and sparse walls, respectively, which were mainly identified as persistent-defined fibrosis. Notably, the size of the honeycomb cysts did not appear to affect lifetime. In contrast, the sparse wall surrounding the cysts was associated with a relatively shorter (moderate) lifetime of H_2 cycles, and pure GGO without reticulations was associated with a very short lifetime. GGOs are divided into GGOs superimposed by a reticular pattern (simply termed reticulation) and pure GGOs, and the latter is not a feature of IPF (20). Because GGO is considered an early stage of fibrosis in interstitial lung disease associated with systemic sclerosis (32), a future study should define GGO using persistent homology and test whether GGO precedes the honeycomb region in fibrotic lung diseases not limited to IPF.

Importantly, the finding that the lifetime substantially differs among GGOs, cysts with sparse walls, and cysts with dense walls suggests that local fibrotic progression could be captured by dynamic changes in the lifetime of H_2 cycles when longitudinal data are available. Furthermore, the mathematical definition of morphological changes would enable establishing a mathematical disease model to simulate disease progression, which is difficult to achieve with experts and DL-based segmentations. We believe that this advantage of persistent homology-based evaluation in conjunction with three-dimensional modeling may uncover the morphological pattern in the local progression of fibrotic lung diseases in future studies.

The emphysema regions were defined as voxels whose neighboring voxels contained more dimension 0 (H_0) cycles with small births and medium lifetimes. This finding is also visualized in Fig. 6, in which the H_0 cycles represent regions with higher density voxels surrounded by lower density voxels. Previous studies have shown that new regions of emphysema develop near preexisting regions of emphysema more frequently than other regions far from preexisting emphysema and can induce the coalescence of two neighboring emphysema regions to cause larger emphysema clusters (33, 34). In other words, new regions of emphysema are less likely to develop randomly in an isolated form (35). Therefore, the increased H_0 cycles with lower birth in emphysema regions might reflect that many normal regions could be isolated by surrounding emphysema regions that expand and form a network over time. Furthermore, because a recently introduced nonrigid registration of paired inspiratory and expiratory CT has allowed identifying emphysema as lower density regions with gas trapping, which is a main functional impairment in emphysema (36, 37), whether persistent homology-based emphysema is associated with emphysematous gas trapping on registered inspiratory and expiratory CT should be performed in a future study.

We performed DL-based segmentation according to a previous paper by Kaji et al. (22) that provides a code for generic image translation algorithms, including the widely used U-Net (23). Because no previous paper has evaluated the model performances of U-Net for this specific segmentation problem of emphysema and fibrosis, we evaluated the performance of the present DL model using manual segmentation as the reference. As shown in Table 2, DL-based fibrosis% and emphysema% were closely associated with fibrosis% and emphysema % measured on the reference segmentation, and the Dice coefficients between the DL-based segmentation and the reference segmentation of fibrosis and emphysema were comparable to those between the independent analyst's manual segmentation and the reference segmentation representing the intero-

bserver variability. These results suggest the validity of the present DL-based segmentation methods.

The results of segmentation by persistent homology-based and DL-based methods were not fully consistent. Persistent homology-based segmentation assigned relatively small cysts surrounded by fibrotic regions as fibrosis, whereas the deep learning method did not assign these cysts as fibrosis (Fig. 4). Differentiating emphysema from honeycomb cysts on CT in fibrotic lungs is challenging in cases with fibrotic lung diseases, and large interobserver variation exists to identify honeycomb cysts on visual inspection (7, 38). Although the present data confirmed the clinical validity of persistent homology-defined fibrosis and emphysema by showing the associations of PH-fibrosis% and PH-emphysema% with D_{LCO} and the CPI and that of PH-fibrosis% with mortality, future studies should be performed to compare the persistent homology-based segmentation of emphysema and fibrosis to corresponding lung histology and explore pathological changes that persistent homology-based emphysema and fibrosis can reflect.

Some limitations are worth mentioning. First, all the CT scans used in this study were reconstructed using a sharp kernel algorithm developed by a single vendor (Canon Medical). The sharp reconstruction kernel increases image contrast, generating more voxels at CT values less than −1,000 HU than the soft reconstruction kernel (39). This phenomenon might affect the present finding that the lower limit for the birth of fibrosis cycles was −1,60 HU. Therefore, whether the proposed method with the same set of parameters can be generalized to CT images obtained with different conditions remains to be elucidated. However, because the definition of each characteristic cycle involves only five parameters, namely, the dimension and the lower and upper limits for birth and lifetime, the present methods can be easily adjusted to analyze CT images reconstructed using a soft kernel algorithm and/or an algorithm from a different vendor. The dimension is fixed, and the other four values have clear meanings as CT values and would not change drastically. These four tunable parameters can be easily determined and used for calibration for different CT configurations. This is another advantage over DL-based segmentation methods. Second, most of the subjects in the two datasets were male. Whether the present findings could be applied to female subjects should be carefully considered.

In conclusion, this study showed that persistent homology could be applied to automatically localize major radiological abnormalities in lung diseases, such as fibrosis and emphysema, on volumetric CT. The data suggest that the lifetime of void components on the persistence diagrams reflects morphological variability in fibrotic changes in IPF, a finding that likely aids in mathematically defining disease progression patterns in future longitudinal studies. Therefore, the persistent homology-based radiological definition of various pathological changes is promising for reproducible, quantitative, and interpretable assessments of the complex structural alterations of chronic lung diseases.

APPENDIX

Persistent Homology Computation

Homological features of a given volumetric CT data were extracted by persistent homology using the open-source program Cubical Ripser (https://github.com/shizuo-kaji/CubicalRipser_3dim) (26). For a CT image, the upper star-filtered cubical complex was built using the CT value; that is, a cubical grid was constructed whose vertexes corresponded to the voxels of CT. Each cell (vertex, edge, square, or cube) is assigned the minimum CT value of its constituent. The union of all the cells with values greater than a threshold value **t** is denoted by $\mathbf{X_t}$. Then, $\cdots \subset \mathbf{X_{t+1}} \subset \mathbf{X_t} \subset \mathbf{X_{t-1}} \subset \cdots$ forms a filtered cell complex. The persistent homology of this complex was computed in the present study. Another popular method to create a filtered complex from an image is to take the distance transform of the binarized image. In the latter, scale-dependent features are mainly captured. In this study, the former method was employed because the CT values have clinical meaning.

Decision Tree Models to Classify the Persistence Diagram of ROIs

Decision tree models were trained to classify ROIs (normal, fibrosis, and emphysema) based on persistence diagrams. To achieve a simple and interpretable definition of the characteristic cycles for fibrosis and emphysema, the number of tree levels was restricted to two. The variable was assumed to have the form "the density of cycles satisfying (1) dimension = d (2) $a < $ birth $< b$ (3) $p < $ lifetime $< q$," where d, a, b, p, and q were parameters and lifetime was the difference between birth and death. These parameters were determined by an automatic grid search to achieve the highest classification accuracy.

Computation Requirement and Time

The analyses based on decision tree models, PCA, and Dice coefficients were performed using the implementation of scikit-learn (40). Computation of the cycle density and its visualization were performed by a custom-made Python script and CuPy (41) with acceleration by a GPU (Nvidia GeForce RTX 2080). For the learning phase, determining characteristic cycles by a grid search to establish the decision tree took approximately an hour for the persistent homology-based method, whereas training the DL-model using the labeled images took a few hours. To process the volume CT data of a subject, computation of the persistent homology and density of the characteristic cycles took ∼20 min, whereas the DL-based method took a few minutes to produce segmentation using the established trained model.

DISCLOSURES

K. Tanizawa worked for the Department of Respiratory Care and Sleep Control Medicine at Kyoto University, which was

REFERENCES

1. **Weibel ER**, **Sapoval B**, **Filoche M.** Design of peripheral airways for efficient gas exchange. *Respir Physiol Neurobiol* 148: 3–21, 2005. doi:10.1016/j.resp.2005.03.005.

2. **Murray CJ**, **Lopez AD.** Measuring the global burden of disease. *N Engl J Med* 369: 448–457, 2013. doi:10.1056/NEJMra1201534.

3. **Raghu G**, **Collard HR**, **Egan JJ**, **Martinez FJ**, **Behr J**, **Brown KK**, et al. An official ATS/ERS/JRS/ALAT statement: Idiopathic pulmonary fibrosis: evidence-based guidelines for diagnosis and management. *Am J Respir Crit Care Med* 183: 788–824, 2011. doi:10.1164/rccm.2009-040GL.

4. **Martinez FJ**, **Safrin S**, **Weycker D**, **Starko KM**, **Bradford WZ**, **King TE Jr**, **Flaherty KR**, **Schwartz DA**, **Noble PW**, **Raghu G**, **Brown KK**, **IPF Study Group.** The clinical course of patients with idiopathic pulmonary fibrosis. *Ann Intern Med* 142: 963–967, 2005. doi:10.7326/0003-4819-142-12_Part_1-200506210-00005.

5. **Vogelmeier CF**, **Criner GJ**, **Martinez FJ**, **Anzueto A**, **Barnes PJ**, **Bourbeau J**, **Celli BR**, **Chen R**, **Decramer M**, **Fabbri LM**, **Frith P**, **Halpin DM**, **Lopez Varela MV**, **Nishimura M**, **Roche N**, **Rodriguez-Roisin R**, **Sin DD**, **Singh D**, **Stockley R**, **Vestbo J**, **Wedzicha JA**, **Agusti A.** Global strategy for the diagnosis, management, and prevention of chronic obstructive lung disease 2017 report: GOLD executive summary. *Eur Respir J* 49: 1700214, 2017. doi:10.1183/13993003.00214-2017.

6. **Lynch DA**, **Austin JH**, **Hogg JC**, **Grenier PA**, **Kauczor HU**, **Bankier AA**, **Barr RG**, **Colby TV**, **Galvin JR**, **Gevenois PA**, **Coxson HO**, **Hoffman EA**, **Newell JD Jr**, **Pistolesi M**, **Silverman EK**, **Crapo JD.** CT-definable subtypes of chronic obstructive pulmonary disease: a statement of the Fleischner Society. *Radiology* 277: 192–205, 2015. doi:10.1148/radiol.2015141579.

7. **Lynch DA**, **Sverzellati N**, **Travis WD**, **Brown KK**, **Colby TV**, **Galvin JR**, **Goldin JG**, **Hansell DM**, **Inoue Y**, **Johkoh T**, **Nicholson AG**, **Knight SL**, **Raoof S**, **Richeldi L**, **Ryerson CJ**, **Ryu JH**, **Au W.** Diagnostic criteria for idiopathic pulmonary fibrosis: A Fleischner Society White Paper. *Lancet Respir Med* 6: 138–153, 2018. doi:10.1016/S2213-2600(17)30433-2.

8. **Verleden SE**, **Tanabe N**, **McDonough JE**, **Vasilescu DM**, **Xu F**, **Wuyts WA**, **Piloni D**, **De Sadeleer L**, **Willems S**, **Mai C**, **Hostens J**, **Cooper JD**, **Verbeken EK**, **Verschakelen J**, **Galban CJ**, **Van Raemdonck DE**, **Colby TV**, **Decramer M**, **Verleden GM**, **Kaminski N**, **Hackett TL**, **Vanaudenaerde BM**, **Hogg JC.** Small airways pathology in idiopathic pulmonary fibrosis: A retrospective cohort study. *Lancet Respir Med* 8: 573–584, 2020. doi:10.1016/S2213-2600(19)30356-X.

9. **Cottin V**, **Nunes H**, **Brillet PY**, **Delaval P**, **Devouassoux G**, **Tillie-Leblond I**, **Israel-Biet D**, **Court-Fortune I**, **Valeyre D**, **Cordier JF**, **Groupe d'Etude et de Recherche sur les Maladies Orphelines Pulmonaires (GERM O P).** Combined pulmonary fibrosis and emphysema: a distinct underrecognised entity. *Eur Respir J* 26: 586–593, 2005. doi:10.1183/09031936.05.00021005.

10. **Jacob J**, **Bartholmai BJ**, **Rajagopalan S**, **Kokosi M**, **Maher TM**, **Nair A**, **Karwoski R**, **Renzoni E**, **Walsh SLF**, **Hansell DM**, **Wells AU.** Functional and prognostic effects when emphysema complicates idiopathic pulmonary fibrosis. *Eur Respir J* 50: 1700379, 2017. doi:10.1183/13993003.00379-2017.

11. **Jacob J**, **Bartholmai BJ**, **Rajagopalan S**, **van Moorsel CHM**, **van Es HW**, **van Beek FT**, **Struik MH**, **Kokosi M**, **Egashira R**, **Brun AL**, **Nair A**, **Walsh SLF**, **Cross G**, **Barnett J**, **de Lauretis A**, **Judge EP**, **Desai S**, **Karwoski R**, **Ourselin S**, **Renzoni E**, **Maher TM**, **Altmann A**, **Wells AU.** Predicting outcomes in idiopathic pulmonary fibrosis using automated computed tomographic analysis. *Am J Respir Crit Care Med* 198: 767–776, 2018. doi:10.1164/rccm.201711-2174OC.

12. **Jacob J**, **Bartholmai BJ**, **Rajagopalan S**, **Karwoski R**, **Nair A**, **Walsh SLF**, **Barnett J**, **Cross G**, **Judge EP**, **Kokosi M**, **Renzoni E**, **Maher TM**, **Wells AU.** Likelihood of pulmonary hypertension in patients with idiopathic pulmonary fibrosis and emphysema. *Respirology* 23: 593–599, 2018. doi:10.1111/resp.13231.

13. **Lai M.** Deep learning for medical image segmentation (Preprint). *arXiv* 1505.02000, 2015.

14. **Edelsbrunner H**, **Letscher D**, **Zomorodian A.** Topological persistence and simplification. In: *Proceedings 41st Annual Symposium on Foundations of Computer Science*, 2000, p. 454–463.

15. **Adams H**, **Chepushtanova S**, **Emerson T**, **Hanson E**, **Kirby M**, **Motta F**, **Neville R**, **Peterson C**, **Shipman P**, **Ziegelmeier L.** Persistence images: a stable vector representation of persistent homology (Preprint). *arXiv* 1507.06217, 2015.

16. **Hiraoka Y**, **Nakamura T**, **Hirata A**, **Escolar EG**, **Matsue K**, **Nishiura Y.** Hierarchical structures of amorphous solids characterized by persistent homology. *Proc Natl Acad Sci U S A* 113: 7035–7040, 2016. doi:10.1073/pnas.1520877113.

17. **Belchi F**, **Pirashvili M**, **Conway J**, **Bennett M**, **Djukanovic R**, **Brodzki J.** Lung topology characteristics in patients with chronic obstructive pulmonary disease. *Sci Rep* 8: 5341, 2018. doi:10.1038/s41598-018-23424-0.

18. **Adcock A**, **Rubin D**, **Carlsson G.** Classification of hepatic lesions using the matching metric. *Comput Vis Image Underst* 121: 36–42, 2014. doi:10.1016/j.cviu.2013.10.014.

19. **Oyama A**, **Hiraoka Y**, **Obayashi I**, **Saikawa Y**, **Furui S**, **Shiraishi K**, **Kumagai S**, **Hayashi T**, **Kotoku J.** Hepatic tumor classification using texture and topology analysis of non-contrast-enhanced three-dimensional T1-weighted MR images with a radiomics approach. *Sci Rep* 9: 8764, 2019. doi:10.1038/s41598-019-45283-z.

20. **Raghu G**, **Remy-Jardin M**, **Myers JL**, **Richeldi L**, **Ryerson CJ**, **Lederer DJ**, et al. Diagnosis of idiopathic pulmonary fibrosis. An official ATS/ERS/JRS/ALAT clinical practice guideline. *Am J Respir Crit Care Med* 198: e44–e68, 2018. doi:10.1164/rccm.201807-1255ST.

21. **Wells AU**, **Desai SR**, **Rubens MB**, **Goh NS**, **Cramer D**, **Nicholson AG**, **Colby TV**, **Du Bois RM**, **Hansell DM.** Idiopathic pulmonary fibrosis: a composite physiologic index derived from disease extent observed by computed tomography. *Am J Respir Crit Care Med* 167: 962–969, 2003. doi:10.1164/rccm.2111053.

22. **Kaji S**, **Kida S.** Overview of image-to-image translation by use of deep neural networks: denoising, super-resolution, modality conversion, and reconstruction in medical imaging. *Radiol Phys Technol* 12: 235–248, 2019. doi:10.1007/s12194-019-00520-y.

23. **Ronneberger O**, **Fischer P**, **Brox T.** CU-Net: convolutional networks for biomedical image segmentation (Preprint). *arXiv* 1505.04597, 2015.

24. **Tanizawa K**, **Handa T**, **Nagai S**, **Hirai T**, **Kubo T**, **Oguma T**, **Ito I**, **Ito Y**, **Watanabe K**, **Aihara K**, **Ikezoe K**, **Oga T**, **Chin K**, **Izumi T**, **Mishima M.** Clinical impact of high-attenuation and cystic areas on computed tomography in fibrotic idiopathic interstitial pneumonias. *BMC Pulm Med* 15: 74, 2015. doi:10.1186/s12890-015-0069-0.

25. **Tanabe N**, **Vasilescu DM**, **Kirby M**, **Coxson HO**, **Verleden SE**, **Vanaudenaerde BM**, **Kinose D**, **Nakano Y**, **Pare PD**, **Hogg JC.** Analysis of airway pathology in COPD using a combination of computed tomography, micro-computed tomography and histology. *Eur Respir J* 51: 1701245, 2018. doi:10.1183/13993003.01245-2017.

26. **Kaji S**, **Sudo T.** Cubical Ripser: Software for computing persistent homology of image and volume data (Preprint). *arXiv* 2005.12692, 2020.

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

28. **Rao CR.** The utilization of multiple measurements in problems of biological classification. *J R Stat Soc: B (Methodol)* 10: 159–193, 1948. doi:10.1111/j.2517-6161.1948.tb00008.x.

29. **R Core Team.** R: A Language And Environment For Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing, 2016. http://www.r-project.org/.

30. **Dunn OJ.** Multiple comparisons using rank sums. *Technometrics* 6: 241–252, 1964. doi:10.1080/00401706.1964.10490181.

31. **Perea JA**, **Munch E**, **Khasawneh FA.** Approximating continuous functions on persistence diagrams using template functions (Preprint). *arXiv* 1902.07190, 2019.

32. **Launay D**, **Remy-Jardin M**, **Michon-Pasturel U**, **Mastora I**, **Hachulla E**, **Lambert M**, **Delannoy V**, **Queyrel V**, **Duhamel A**, **Matran R**, **De Groote P**, **Hatron PY.** High resolution computed tomography in fibrosing alveolitis associated with systemic sclerosis. *J Rheumatol* 33: 1789–1801, 2006.

33. **Mondonedo JR**, **Sato S**, **Oguma T**, **Muro S**, **Sonnenberg AH**, **Zeldich D**, **Parameswaran H**, **Hirai T**, **Suki B.** CT imaging-based low-attenuation super clusters in three dimensions and the progression of emphysema. *Chest* 155: 79–87, 2019. doi:10.1016/j.chest.2018.09.014.

34. **Tanabe N**, **Muro S**, **Hirai T**, **Oguma T**, **Terada K**, **Marumo S**, **Kinose D**, **Ogawa E**, **Hoshino Y**, **Mishima M.** Impact of exacerbations on emphysema progression in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 183: 1653–1659, 2011. doi:10.1164/rccm.201009-1535OC.

35. **Suki B**, **Lutchen KR**, **Ingenito EP.** On the progressive nature of emphysema: roles of proteases, inflammation, and mechanical forces. *Am J Respir Crit Care Med* 168: 516–521, 2003. doi:10.1164/rccm.200208-908PP.

36. **Galban CJ**, **Han MK**, **Boes JL**, **Chughtai KA**, **Meyer CR**, **Johnson TD**, **Galban S**, **Rehemtulla A**, **Kazerooni EA**, **Martinez FJ**, **Ross BD.** Computed tomography-based biomarker provides unique signature for diagnosis of COPD phenotypes and disease progression. *Nat Med* 18: 1711–1715, 2012. doi:10.1038/nm.2971.

37. **Kirby M**, **Yin Y**, **Tschirren J**, **Tan WC**, **Leipsic J**, **Hague CJ**, **Bourbeau J**, **Sin DD**, **Hogg JC**, **Coxson HO**; **CanCOLD Collaborative Research Group and the Canadian Respiratory Research Network.** A novel method of estimating small airway disease using inspiratory-to-expiratory computed tomography. *Respiration* 94: 336–345, 2017. doi:10.1159/000478865.

38. **Watadani T**, **Sakai F**, **Johkoh T**, **Noma S**, **Akira M**, **Fujimoto K**, **Bankier AA**, **Lee KS**, **Muller NL**, **Song JW**, **Park JS**, **Lynch DA**, **Hansell DM**, **Remy-Jardin M**, **Franquet T**, **Sugiyama Y.** Interobserver variability in the CT assessment of honeycombing in the lungs. *Radiology* 266: 936–944, 2013. doi:10.1148/radiol.12112516.

39. **Gierada DS**, **Bierhals AJ**, **Choong CK**, **Bartel ST**, **Ritter JH**, **Das NA**, **Hong C**, **Pilgram TK**, **Bae KT**, **Whiting BR**, **Woods JC**, **Hogg JC**, **Lutey BA**, **Battafarano RJ**, **Cooper JD**, **Meyers BF**, **Patterson GA.** Effects of CT section thickness and reconstruction kernel on emphysema quantification relationship to the magnitude of the CT emphysema index. *Acad Radiol* 17: 146–156, 2010. doi:10.1016/j.acra.2009.08.007.

40. **Pedregosa F**, **Varoquaux G**, **Gramfort A**, **Michel V**, **Thirion B**, **Grisel O**, **Blondel M**, **Müller A**, **Nothman J**, **Louppe G**, **Prettenhofer P**, **Weiss R**, **Dubourg V**, **Vanderplas J**, **Passos A**, **Cournapeau D**, **Brucher M**, **Perrot M**, **Duchesnay É.** Scikit-learn: machine learning in Python (Preprint). *arXiv* 1201.0490, 2012.

41. **Okuta R**, **Unno Y**, **Nishino D**, **Hido S**, **Cupy LC.** A numpy-compatible library for nvidia gpu calculations. In: *Proceedings of Workshop on Machine Learning Systems (LearningSys) in The Thirty-first Annual Conference on Neural Information Processing Systems (NIPS)*, 2017.