

# PleThora: Pleural effusion and thoracic cavity segmentations in diseased lungs for benchmarking chest CT processing pipelines

Kendall J. Kiser<sup>a)</sup>

*John P. and Kathrine G. McGovern Medical School, Houston, TX, USA*

*Center for Precision Health, UTHealth School of Biomedical Informatics, Houston, TX, USA*

*Department of Radiation Oncology, University of Texas MD Anderson Cancer Center, Houston, TX, USA*

Sara Ahmed and Sonja Stieb

*Department of Radiation Oncology, University of Texas MD Anderson Cancer Center, Houston, TX, USA*

Abdallah S. R. Mohamed

*Department of Radiation Oncology, University of Texas MD Anderson Cancer Center, Houston, TX, USA*

*MD Anderson Cancer Center-UTHealth Graduate School of Biomedical Sciences, Houston, TX, USA*

Hesham Elhalawani

*Department of Radiation Oncology, Cleveland Clinic Taussig Cancer Center, Cleveland, OH, USA*

Peter Y. S. Park, Nathan S. Doyle, and Brandon J. Wang

*Department of Diagnostic and Interventional Imaging, John P. and Kathrine G. McGovern Medical School, Houston, TX, USA*

Arko Barman, Zhao Li, and W. Jim Zheng

*Center for Precision Health, UTHealth School of Biomedical Informatics, Houston, TX, USA*

Clifton D. Fuller<sup>a)</sup>

*Department of Radiation Oncology, University of Texas MD Anderson Cancer Center, Houston, TX, USA*

*MD Anderson Cancer Center-UTHealth Graduate School of Biomedical Sciences, Houston, TX, USA*

Luca Giancardo<sup>a)</sup>

*Center for Precision Health, UTHealth School of Biomedical Informatics, Houston, TX, USA*

*Department of Radiation Oncology, Cleveland Clinic Taussig Cancer Center, Cleveland, OH, USA*

(Received 7 April 2020; revised 22 July 2020; accepted for publication 27 July 2020; published 28 August 2020)

This manuscript describes a dataset of thoracic cavity segmentations and discrete pleural effusion segmentations we have annotated on 402 computed tomography (CT) scans acquired from patients with non-small cell lung cancer. The segmentation of these anatomic regions precedes fundamental tasks in image analysis pipelines such as lung structure segmentation, lesion detection, and radiomics feature extraction. Bilateral thoracic cavity volumes and pleural effusion volumes were manually segmented on CT scans acquired from The Cancer Imaging Archive “NSCLC Radiomics” data collection. Four hundred and two thoracic segmentations were first generated automatically by a U-Net based algorithm trained on chest CTs without cancer, manually corrected by a medical student to include the complete thoracic cavity (normal, pathologic, and atelectatic lung parenchyma, lung hilum, pleural effusion, fibrosis, nodules, tumor, and other anatomic anomalies), and revised by a radiation oncologist or a radiologist. Seventy-eight pleural effusions were manually segmented by a medical student and revised by a radiologist or radiation oncologist. Interobserver agreement between the radiation oncologist and radiologist corrections was acceptable. All expert-vetted segmentations are publicly available in NIfTI format through The Cancer Imaging Archive at <https://doi.org/10.7937/tcia.2020.6c7y-gq39>. Tabular data detailing clinical and technical metadata linked to segmentation cases are also available. Thoracic cavity segmentations will be valuable for developing image analysis pipelines on pathologic lungs — where current automated algorithms struggle most. In conjunction with gross tumor volume segmentations already available from “NSCLC Radiomics,” pleural effusion segmentations may be valuable for investigating radiomics profile differences between effusion and primary tumor or training algorithms to discriminate between them. © 2020 The Authors. *Medical Physics* published by Wiley Periodicals LLC on behalf of American Association of Physicists in Medicine. [<https://doi.org/10.1002/mp.14424>]

**Key words:** computer-aided decision support systems, image segmentation techniques, image processing, informatics in imaging, quantitative imaging

## 1. INTRODUCTION

Automated or semi-automated algorithms aimed at analyzing chest computed tomography (CT) scans typically require the creation of a three-dimensional (3D) map of the volume-of-interest (VOI) as the initial step.<sup>1,2</sup> For example, a common first step for lung tumor detection on CT scans is lung segmentation.<sup>3</sup> Effective strategies for identifying healthy lungs have existed at least since Hu et al. introduced a method for segmenting healthy lung parenchyma based on gray-level thresholding.<sup>4</sup> Nevertheless, identifying the initial VOI in pathologic lungs remains an obstacle.<sup>5,6</sup> Many pathologic states — such as pleural effusion, severe fibrosis, or tumor — can alter the space lung would normally occupy, and in this circumstance the VOI is not only the lung but the thoracic cavity. To build image processing pipelines intended to analyze chest CTs with substantively altered thoracic anatomy, identifying this VOI is critical.

Preprocessing strategies to identify thoracic VOIs in the presence of pathology have been described. In 2014, Mansoor et al. presented a seminal approach that identified CTs with large volumetric differences between autosegmented lung and the thoracic cage and refined these lung segmentations using texture-based features.<sup>7</sup> Subsequent studies have approached VOI identification in myriad ways, such as threshold-based methodologies,<sup>8</sup> deep learning architectures,<sup>9–13</sup> anatomic or shape-prior models,<sup>5,14–16</sup> and region-growing methods.<sup>17,18</sup>

As methodologies to identify thoracic VOIs in pathologic lungs march forward, data to train and vet them must keep pace. Hofmanninger et al. cited a paucity of diverse data — not inferior methodologies — as the principal obstacle in pathologic lung segmentation.<sup>19</sup> Echoing this point, Shaukat et al. observed that automated lung nodule detection system optimization is usually limited to just one dataset.<sup>20</sup> Many datasets consist exclusively of diseases typified by mild to moderate anatomic change — for example, chronic obstructive pulmonary disorder or interstitial lung disease.<sup>10,11</sup> However, disease processes commonly beset by lung consolidations, effusions, masses, and other opacities that dramatically alter the thoracic VOI are radiographically distinct.

We present PleThora, a dataset of *pleural* effusion and left and right *thoracic* cavity segmentations delineated on 402 CT scans from The Cancer Imaging Archive<sup>21</sup> (TCIA) NSCLC-Radiomics collection.<sup>22,23</sup> Many of these cases have dramatic anatomic changes secondary to cancer. Thoracic segmentations include lung parenchyma, primary tumor, atelectasis, adhesions, effusion, and other anatomic variations when present. On scans where effusion is present, separate segmentations labeling pleural effusion alone are also provided. These may serve particular use for correlating effusion radiomics features with clinical data (available in the NSCLC-Radiomics spreadsheet, “Radiomics Lung1.clinical-version3-Oct 2019.csv”) or studying how these features differ between effusion and primary tumor. Thoracic segmentations were generated automatically by a U-Net based deep learning algorithm trained on lungs without cancer, manually

corrected by a medical student, and revised by a radiation oncologist or a radiologist. Pleural effusion segmentations were manually delineated by a medical student and revised by a radiologist. Expert gross tumor volume (GTV) segmentations already provided by the NSCLC-Radiomics collection informed our segmentations and made possible delineation of pleural effusion volumes that excluded GTV.

## 2. ACQUISITION AND VALIDATION METHODS

### 2.A. Segmentation acquisition

Four hundred twenty-two Digital Imaging and Communications in Medicine (DICOM) chest CT datasets and 318 corresponding “RTSTRUCT” DICOM segmentations (featuring GTVs) were downloaded from the TCIA NSCLC-Radiomics collection in January 2019. Four hundred and two CT scans were successfully converted from DICOM to Neuroimaging Informatics Technology Initiative (NIfTI) format using a free executable called “dcm2nii”.<sup>24,25</sup> These 402 scans comprise the dataset upon which our thoracic cavity and pleural effusion segmentations were delineated.

#### 2.A.1. Thoracic cavity segmentation acquisition

After converting each CT dataset to NIfTI format, lungs were automatically segmented by a publicly available 3D U-Net lung segmentation algorithm<sup>26</sup> that had been trained on a private dataset of approximately 200 chest CTs acquired from patients without cancer. Lung segmented reasonably well in subjects with minimal anatomic variation due to disease, but poorly in subjects with significant disease. A fourth-year medical student manually expanded each segmentation to include the thoracic cavity volume normally occupied by healthy lung. Among other anatomic anomalies, the expansion included normal, pathologic, and atelectatic lung parenchyma, pleural effusion, fibrosis, nodules, tumor, and compensatory anatomic variants such as enlarged collateral vessels. It also included major hilar vessels and bronchi that are not always segmented as part of the lung but are a frequent site of pathologic change. Importantly, effort was made to include all primary tumors in the segmentation even if this invaded the mediastinum or extended out of the thoracic cage. “RTSTRUCT” GTV segmentations from the NSCLC-Radiomics collection assisted in this determination. Nodal disease in the mediastinum was not targeted. Each hemithorax, right and left, was segmented as a separate structure but saved under the same segmentation file.

The medical student’s segmentations passed to at least one physician reviewer — either a radiologist or a radiation oncologist — to be vetted. As necessary, reviewers expanded or contracted the segmentations to include the target thoracic volumes. The most recent “RTSTRUCT” GTV segmentations (which were updated by the NSCLC-Radiomics dataset authors in October 2019) were available to reviewers for reference as necessary. The thoracic cavity segmentations we make public are vetted segmentations. The thoracic cavity

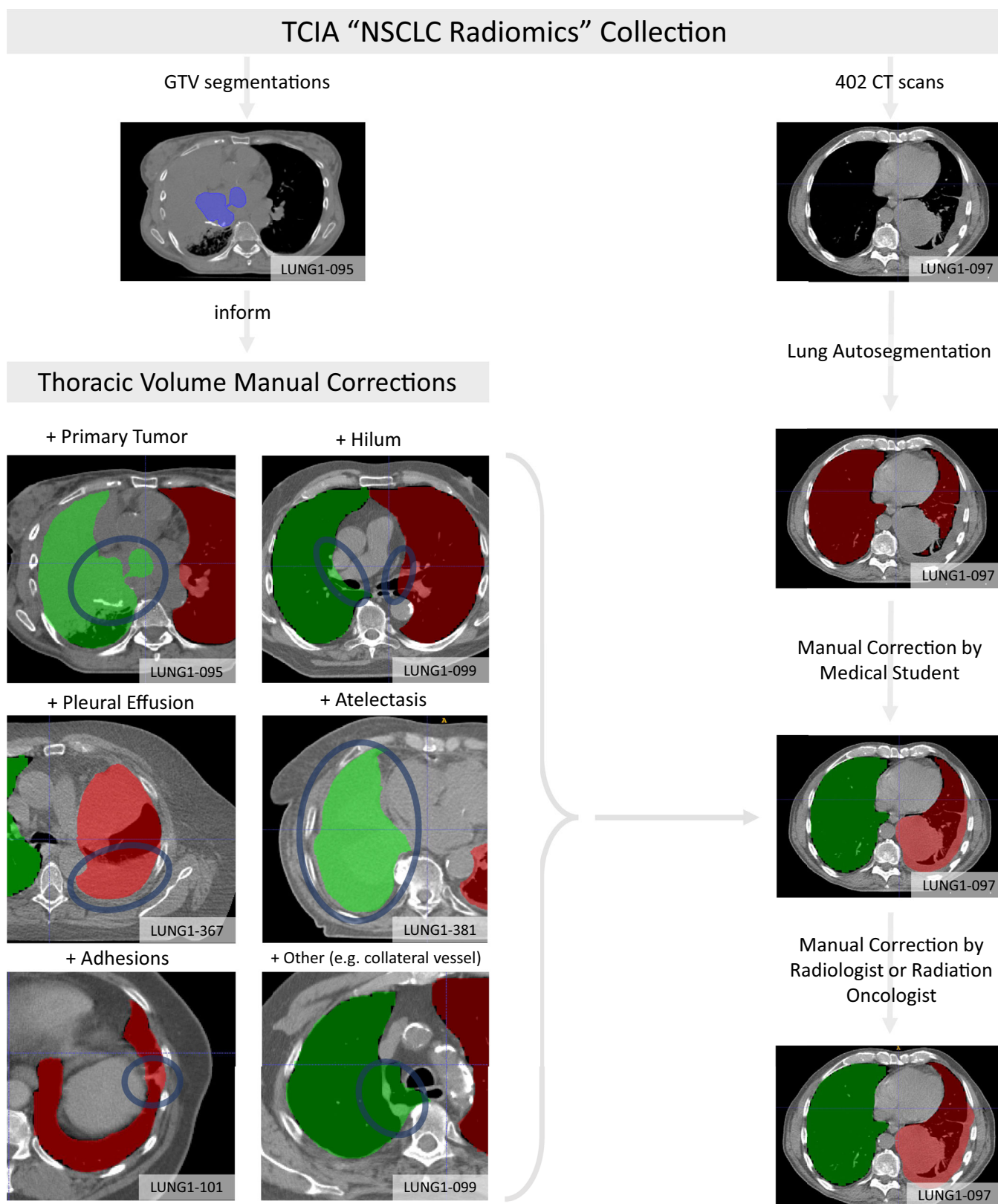


FIG. 1. Thoracic cavity volumes were segmented automatically then iteratively corrected by a medical student and at least one radiologist or radiation oncologist to include the entire hemithoraces. [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

segmentation workflow is illustrated in Fig. 1 with examples of various lesions that were included in the delineations.

### 2.A.2. Pleural effusion segmentation acquisition

A fourth-year medical student identified a subset of 78 CT scans with trace to massive pleural effusions. The medical student segmented effusions de novo rather than from an autosegmented prior (unlike the thoracic cavity segmentations). Spaces where primary tumor overlapped with effusion were excluded from the segmentation, and NSCLC-Radiomics GTV segmentations determined the effusion segmentation extent (exemplified in Fig. 2). In most subjects, an effusion in a single hemithorax was segmented. However,

effusions were bilateral in 19 subjects, and in this circumstance both sides were segmented and saved under a single structure label.

The medical student's pleural effusion segmentations were vetted by at least two physicians and corrected as necessary. Like the thoracic cavity segmentations, the pleural effusion segmentations we make public are vetted by physicians. The pleural effusion segmentation workflow is illustrated in Fig. 2.

### 2.B. Computational tools

Manual segmentations were delineated using ITK-SNAP v.3.6.<sup>27</sup> ITK-SNAP provides tools for manual or semi-automated segmentation delineation including segmentation

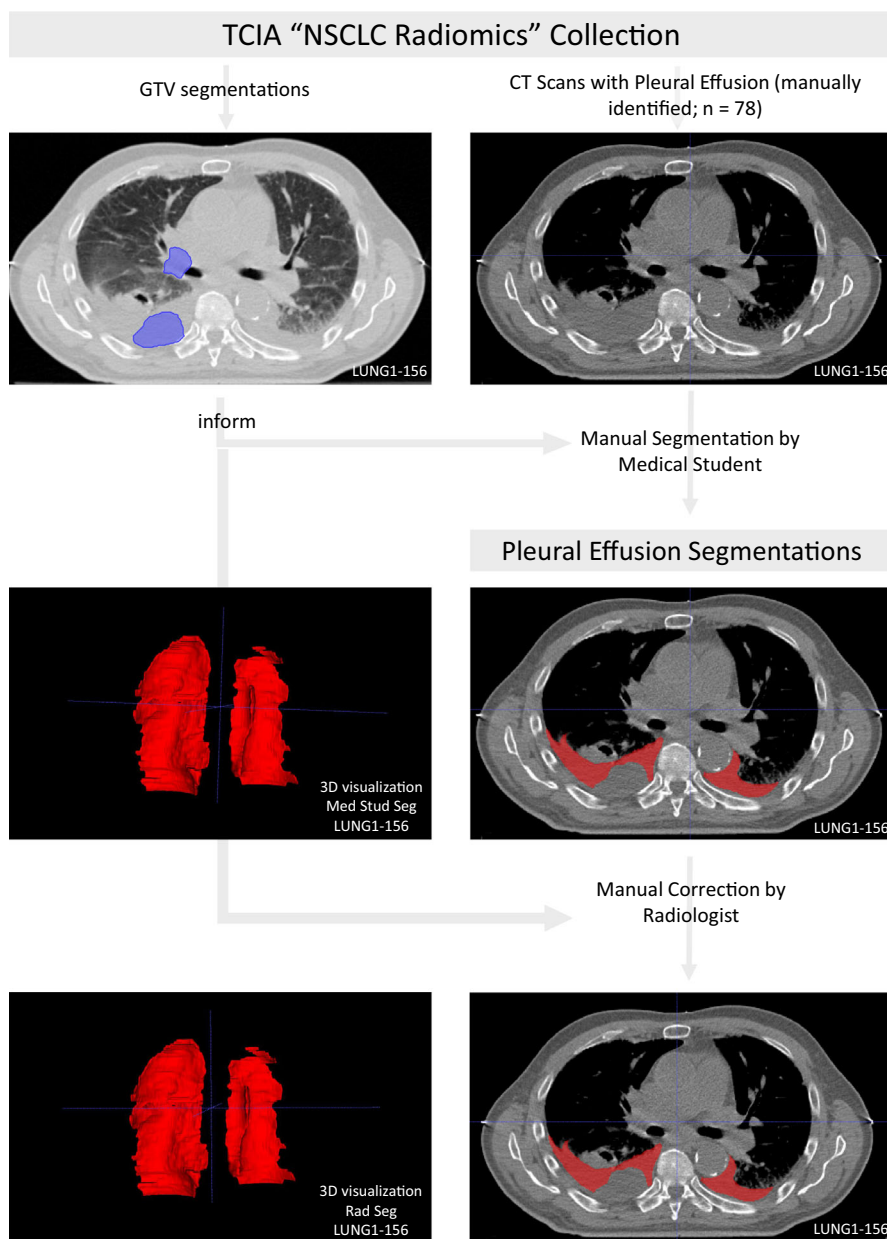


FIG. 2. Pleural effusion segmentations, excluding gross tumor volume, were delineated by a medical student and subsequently corrected by at least two radiologists. [Color figure can be viewed at wileyonlinelibrary.com]



interpolation, 3D segmentation visualization and manipulation, and image contrast adjustment permitting mediastinal and lung window visualization. However, ITK-SNAP does not support structure sets in DICOM-RT format and could not open “RTSTRUCT” files to view GTV segmentations. Therefore, we used a free DICOM-RT viewer called Dicompyler<sup>28</sup> to reference GTV structures while segmenting.

Several additional computational tools transformed CT scans from DICOM to NIfTI format, autosegmented initial thoracic volumes, analyzed segmentations, and organized and visualized metadata. As previously noted, an executable named “dcm2niix”<sup>24</sup> converted CTs from DICOM to NIfTI format and a 3D U-Net<sup>26</sup> automatically generated bilateral lung segmentations. All data were prepared using custom Python scripts leveraging common scientific libraries: NumPy v.1.16.2,<sup>29</sup> Nilearn v.0.5.0,<sup>30</sup> Pandas v.0.23.4,<sup>31</sup> Surface\_distance v.0.1,<sup>32</sup> Scikit-learn v.0.20.3,<sup>33</sup> Matplotlib v.3.0.1,<sup>34</sup> and Seaborn v.0.9.0.<sup>35</sup>

## 2.C. Segmentation quality assessment

To assess the quality of segmentations with respect to interobserver variability, volumetric (i.e., traditional) Dice similarity coefficients<sup>36</sup> (DSC), Cohen’s kappa coefficients,<sup>37</sup> surface Dice similarity coefficients<sup>32</sup> (sDSC), 95th percentile Hausdorff distances<sup>38</sup> (95HD), and symmetric average surface distances (ASD) were computed between expert reviewer segmentations. These metrics are discussed in the following subsections, and lung segmentation interobserver variability reference values are given where possible. Although methodologies to automate CT pleural effusion segmentation<sup>39–42</sup> or decrease qualitative size estimation variability<sup>43</sup> have been described, we found only one limited report of pleural effusion segmentation interobserver variability<sup>39</sup> against which we can compare our results (despite extensive PubMed searches).

### 2.C.1. Volumetric Dice similarity coefficient

The DSC measures interobserver agreement and ranges from 0 to 1, where 0 indicates no agreement and 1 indicates perfect agreement. In the context of image segmentation, the DSC relates the overlap between two segmentations to their total volumes; mathematically, this is twice the shared volume divided by the sum of their total volumes. No consensus dictates what constitutes a “good” DSC because the DSC is sensitive to the volume of the target structure. The American Association of Physicists in Medicine (AAPM) Task Group 132 notes that the contouring uncertainty of a structure can be expected to be a DSC of 0.8–0.9, while cautioning that “very large or very small structures may have different expected DSC values for contour uncertainty.”<sup>44</sup> For lung segmentation, DSC interobserver variability between three medical physicist organizers of the 2017 AAPM Thoracic Auto-Segmentation Challenge was reported as  $0.956 \pm 0.019$  and  $0.955 \pm 0.019$  for the left and right lungs, respectively.<sup>45</sup> For pleural effusion segmentation, Yao et al.<sup>39</sup> reported mean DSC observer variability between a research fellow, the same

fellow months later, and an image processing technologist to be about 0.73. Comparisons against this reference value should be made cautiously because it was calculated for only 12 CT scans with unspecified pleural effusion volumes.

### 2.C.2. Cohen’s Kappa coefficient

The Cohen’s kappa ( $\kappa$ ) coefficient measures interobserver reliability for qualitative observations with mutually exclusive classifications. We computed  $\kappa$  between expert reviewer segmentations by treating each voxel as a qualitative datum. We transformed the segmentations to numerical arrays where each array value assumed one of two binary values depending on whether the reviewer included the voxel as part of the target structure.  $\kappa$  is similar to the DSC in its computation of interobserver agreement, but also assesses a likelihood of chance agreement and penalizes accordingly. Like the DSC,  $\kappa$  generally ranges between 0 and 1, although a result  $< 0$  is possible. Results  $> 0.6$  are generally considered good and greater than 0.8 are considered very good,<sup>46</sup> although like the DSC,  $\kappa$  inflates and deflates artificially depending on the target’s volume.

### 2.C.3. Surface Dice similarity coefficient

Recognizing the limitations of the traditional DSC with respect to volume, Nikolov et al.<sup>32</sup> introduced a novel way to compute Dice’s coefficient called the surface DSC. As its name suggests, the sDSC’s inputs are segmentation surfaces rather than volumes. A primary advantage of the sDSC over the traditional DSC is increased robustness to segmentation size variation. The sDSC is not agnostic to size, but it inflates less dramatically with size than the volumetric DSC. The sDSC computation accepts a tolerance parameter whereby differences between two surfaces can be tolerated as the same surface. We set this parameter equal to zero in order to capture all differences between physician segmentations. Because the sDSC is novel, reference values for this metric are not yet widely reported. However, Vaassen et al.<sup>47</sup> calculated the sDSC between automatically generated and radiotherapist-corrected lung segmentations and reported the majority of values to be greater than 0.85. Reasoning that inter-physician agreement ought to be at least as good as autosegmentation-radiotherapist agreement, we suggest 0.85 as an acceptable mean reference value for thoracic cavity segmentations.

### 2.C.4. Hausdorff distance

In contrast to the DSC,  $\kappa$ , and sDSC, which measure fractional overlap between segmentations, HDs are geometric distances between segmentation surfaces. Larger HDs signify worse interobserver agreement. Computing a HD requires determining the minimum distances from every point on the surface of segmentation A to every point on the surface of segmentation B, and the same from B to A, and arranging them in ascending order. We report the 95HD, which is the

value at the 95th percentile of the ordered minimum distances. 95HD interobserver variability reported by the AAPM Thoracic Auto-Segmentation Challenge was  $5.17 \pm 2.73$  and  $6.71 \pm 3.91$  for left and right lung segmentations, respectively.<sup>45</sup>

2.C.5. Average surface distance

The directed ASD from segmentation A to segmentation B is the average of the minimum distances from every point in the surface of A to every point in the surface of B. The directed ASD from B to A is calculated similarly. We report a symmetric ASD value that averages the two directed ASDs. Symmetric ASD interobserver variability reported by the AAPM Thoracic Auto-Segmentation Challenge was

TABLE I. Seven radiologists (Rad) and radiation oncologists (RO) collaborated to review and correct 402 thoracic cavity segmentations and 78 pleural effusion segmentations delineated by a fourth-year medical student.

Expert reviewer	Years of experience
Rad1	4
Rad2	2
Rad3	1
Rad4	3
RO1	4
RO2	11
RO3	5

$1.51 \pm 0.67$  and  $1.87 \pm 0.87$  for left and right lung segmentations, respectively.<sup>45</sup>

2.C.6. Expert reviewers

Expert reviewers consisted of four radiologists and three radiation oncologists with varying years of experience (Table I). All thoracic cavity segmentations were reviewed by at least one expert, and randomly selected, unique subsets were reviewed by two. All pleural effusion segmentations were reviewed by at least two experts and a subset was reviewed by three. Relationships between reviewers are illustrated in Fig. 3. Median and minimum values for each spatial similarity metric for each reviewer pair are given in Table II. Pair-wise Mann-Whitney *U* tests<sup>48</sup> conducted between reviewer pair distributions suggest that they are significantly different from one another for all metrics ( $P < 0.001$ ). The following paragraphs discuss how interobserver variability in our study compares with variability described in other studies.

In general, thoracic cavity segmentation pairs enjoyed a good level of agreement. Gauged by DSC, 95HD, and symmetric ASD spatial similarity metrics, thoracic cavity segmentation interobserver variability was similar to reported interobserver variability for lung segmentation<sup>45</sup> and similar to values achieved by state-of-the-art deep learning architectures trained for lung autosegmentation.<sup>19</sup> Spatial similarity metrics improved with each iteration of corrections. For example, DSC values calculated between initial automated segmentations and substantially corrected medical student segmentations were relatively low ( $DSC_{min} = 0.353$ ,

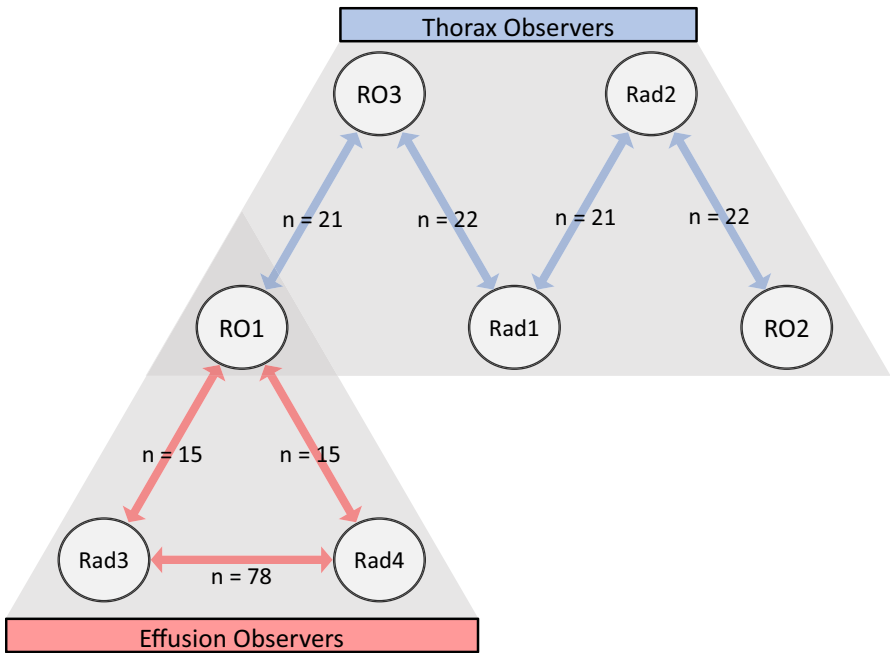


FIG. 3. A schematic of interobserver comparisons, with the number of segmentation cases shared between observer pairs given as *n*. All 78 pleural effusion segmentations were reviewed and as necessary corrected by two radiologists: Rad3 and Rad4. A subset of 15 pleural effusion segmentations were also reviewed and corrected by RO1. In contrast, not all 402 thoracic cavity segmentations were reviewed by two physicians. Rather, four subsets of 21 or 22 thoracic cavity segmentations were randomly selected for dual review. All members of a given subset were exclusive to that subset. [Color figure can be viewed at wileyonlinelibrary.com]

TABLE II. Median and minimum values for Dice similarity coefficient (DSC), surface DSC,  $\kappa$ , 95HD, and symmetric ASD spatial similarity metrics calculated between paired physician segmentations. The distributions for each observer pair are significantly different from one another for all metrics (paired Mann-Whitney  $U$  test<sup>48</sup>  $P < 0.001$ ). However, interobserver variability between pairs of physician-vetted segmentations is generally acceptable. Median DSC, 95HD and symmetric ASD values for thoracic cavity segmentations are comparable to mean interobserver variability values reported by the 2017 AAPM Thoracic Auto-Segmentation Challenge for lung segmentations.<sup>45</sup> In general, pleural effusion segmentation interobserver agreement is also acceptable but more variable, reflecting both variation in pleural effusion size and inclusion or exclusion of trace pleural fluid.

		Pleural effusions			Thoracic cavities			
Metric		RO1-Rad3	RO1-Rad4	Rad3-Rad4	RO1-RO3	Rad1-RO3	Rad1-Rad2	RO2-Rad2
Conformality metrics (unitless)								
DSC	Med	0.81	0.85	0.93	0.99	1.00	1.00	1.00
Min	0.10 <sup>a</sup>	0.26	0.20	0.96	0.91 <sup>a</sup>	0.99	0.97	
sDSC	Med	0.62	0.77	0.87	0.94	0.98	0.98	1.00
Min	0.20 <sup>a</sup>	0.32	0.21	0.73 <sup>a</sup>	0.82	0.94	0.88	
Kappa	Med	0.81	0.85	0.93	0.99	0.99	0.99	0.99
Min	0.10 <sup>a</sup>	0.26	0.20	0.96	0.90 <sup>a</sup>	0.99	0.97	
Surface distance metrics (mm)								
95HD	Med	24.00	21.65	5.31	1.95	0.00	0.00	0.00
Max	127.83	127.80	161.48 <sup>a</sup>	11.35	55.11 <sup>a</sup>	0.98	24.82	
ASD	Med	1.82	2.45	0.79	0.25	0.05	0.03	0.00
Max	23.53	22.39	33.49 <sup>a</sup>	1.01	5.78 <sup>a</sup>	0.12	1.68	

<sup>a</sup>Select cases with extreme spatial similarity metric values are explored visually in Fig. 6.

DSC<sub>med</sub> = 0.958, IQR = 0.041). In contrast, the DSC was never  $<0.9$  between any pair of physician-corrected contours, and the median DSC for each interobserver distribution approximated 0.99 (Fig. 4).

Pleural effusion segmentation interobserver agreement was also consistent, although the distributions of conformality metrics and surface distance metrics generally suggest lower agreement than for thoracic cavity segmentations. Medians DSCs (Fig. 5) compare favorably with Yao et al.'s<sup>39</sup> mean DSC interobserver variability estimate of 0.73, but we again note that the sample in this study was small and the pleural effusion volumes unspecified. Knowing the pleural effusion volume distribution matters because wide variation

in conformality spatial similarity metrics can be partly explained by the spread of pleural effusion volumes. Surface distance metrics can also be influenced by pleural effusion spatial spread (i.e., large distances separating effusion pockets). In a preprint analytic study that uses PleThora,<sup>49</sup> we determined the median right and left pleural effusion volumes to be respectively 58.57 cm<sup>3</sup> (IQR 30.31–113.7 cm<sup>3</sup>) and 50.85 cm<sup>3</sup> (IQR 25.01–142.5 cm<sup>3</sup>). These are an order of magnitude smaller than the mean pleural effusion volumes described in some pleural effusion autosegmentation methodologic studies,<sup>40</sup> which highlights the unique character of this dataset.

Notwithstanding that interobserver agreement was generally very good, a few segmentation cases demonstrated

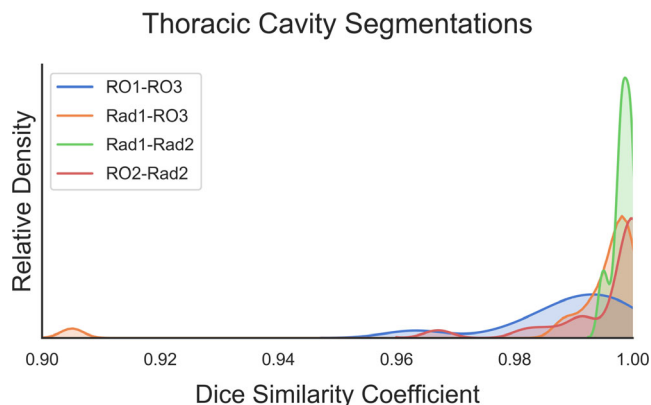


FIG. 4. Dice similarity coefficient distributions reveal consistently strong agreement ( $>0.98$ ) between paired, independently vetted radiologist and radiation oncologist thoracic cavity segmentations. Colored curves are kernel density estimates of DSC distributions. Note that Figs. 4 and 5 do not share the same x axis limits; the difference in DSC distributions in Figs. 4 and 5 is at least partially an artifact of the difference in average volume between thoraces and effusions. [Color figure can be viewed at wileyonlinelibrary.com]

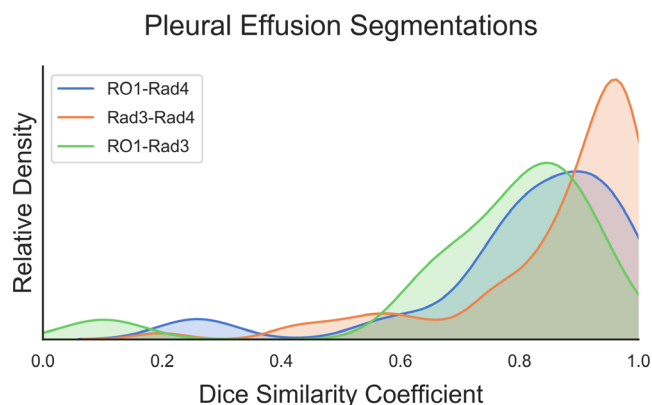


FIG. 5. Dice similarity coefficient (DSC) distributions indicate good agreement ( $>0.8$ ) between most paired, independently vetted radiologist and radiation oncologist pleural effusion segmentations. Interpretation of this result should respect that DSC values calculated on trace pleural effusions are sensitive to variation between segmentations on the order of only a few voxels. As in Fig. 4, colored curves are kernel density estimations of DSC distributions. [Color figure can be viewed at wileyonlinelibrary.com]

markedly poorer agreement than others (Fig. 6). RO1 and Rad3's pleural effusion segmentations for case LUNG1-005 yielded the worst DSC and  $\kappa$  values of any interobserver comparison. Visual inspection reveals a large posterior density that RO1 classified as pleural effusion but Rad3 excluded as an atelectatic lung segment. In this case, Rad4's segmentation arbitrates that the density is not effusion. The segmentation we made publicly available through TCIA in our first version of the dataset was only Rad4's segmentation, but all reviewer segmentations were made available in the second version. Rad3 and Rad4's pleural effusion segmentations for case LUNG1-253 suffered the highest 95HD value of any interobserver comparison. Here, 2D and 3D visual inspection reveals substantial disagreement respecting the superior-inferior and lateral extents of a thin layer of pleural fluid. We did not prospectively define guidelines for whether trace pleural fluid should or should not be segmented, which in hindsight we acknowledge is a limitation of our pleural effusion segmentation methodology. By contrast, the worst discrepancies between thoracic cavity segmentations are easily resolved by our segmentation guidelines. Thoracic cavity segmentations for LUNG1-026 suffered the worst sDSC interobserver variability, which can be attributed to the secondary physician reviewer's erroneous neglect of the full extent of tumor, left hilum, and right pleural effusion. Similarly, the secondary reviewer erroneously excluded a left collapsed lung for case LUNG1-354, which was the thoracic cavity segmentation case with the worst DSC,  $\kappa$ , 95HD, and symmetric ASD. The thoracic cavity segmentations we made publicly available for these cases in the first version of the dataset were the primary reviewer's segmentations, but all reviewers' segmentations were made available in the second version (after correcting the aforementioned errors).

## 2.D. NSCLC-radiomics collection update

The NSCLC-Radiomics collection was updated on 10/23/2019, featuring new "RTSTRUCT" segmentations for all 422 cases, including revised GTVs in some cases. Our thoracic cavity segmentations were reviewed by radiologists or radiation oncologists who had access to "RTSTRUCT" files from the latest collection update. However, these segmentations were first delineated by a medical student at a time when only earlier versions of the "RTSTRUCT" files were available. In contrast, pleural effusion segmentations were all delineated with input from the latest "RTSTRUCT" files.

## 3. DATA FORMAT AND USAGE NOTES

### 3.A. Data and metadata repository

In keeping with findable, accessible, interoperable, re-usable (FAIR) data usage principles,<sup>50</sup> all PleThora thoracic cavity and pleural effusion segmentations have been made available on TCIA at <https://doi.org/10.7937/tcia.2020.6c7y-gq39>.<sup>51</sup> Thoracic cavity segmentations are in a compressed NIfTI format, are named for their respective case and

reviewer (e.g., "LUNG1-001\_thor\_cav\_primary\_reviewer.nii.gz"), and are indexed in folders labeled after their respective NSCLC-Radiomics collection cases (e.g., "LUNG1-001"). Pleural effusion segmentations are likewise saved in a compressed NIfTI format and named for their respective case and reviewer (e.g., "LUNG1-001\_effusion\_first\_reviewer.nii.gz"). Many thoracic cavity and all pleural effusion segmentations were reviewed by two or more experts. In this dataset's original TCIA publication, only primary reviewer segmentations were made available. However, all reviewers' segmentations — primary, secondary, and tertiary — were made available in a recent dataset update (version 2). A spreadsheet entitled "Thorax and Pleural Effusion Segmentation Metadata" contains clinical and technical metadata pertaining to each segmentation or CT scan. It is hosted in the same repository as the segmentations. We strongly recommend users merge it with the NSCLC-Radiomics spreadsheet "Radiomics Lung1.clinical-version3-Oct 2019.csv," which provides ten columns of clinical data tied to each case. We provide Appendix 1 as a data dictionary to clarify the meaning of our spreadsheet's column names.

### 3.B. Baseline for deep learning model performance

To provide a performance baseline for researchers interested in using our thoracic cavity segmentations for deep learning model development, we trained and tested two U-Net models, one based on 2D convolutional layers and one based on 3D convolutional layers. U-Nets are common Convolutional Deep Neural Network architectures and form the basis of many deep learning algorithms for medical image segmentation.<sup>52–56</sup>

CT scans were preprocessed as follows: voxel intensities were clipped to a range of  $[-250, 0]$  Hounsfield units by reassigning voxels less than  $-250$  to  $-250$  and greater than  $0$  to  $0$ , voxels were isotropically resampled to  $1.7$  mm in each dimension, and scans were cropped from the image center to  $256$  by  $256$  by  $128$  voxels. Segmentations were likewise resampled and cropped.

The models were trained using 316 of the 402 primary reviewer segmentations and tested with 86 secondary reviewer segmentations. The latter served as a test dataset because secondary and primary reviews were conducted independently. Nevertheless, because secondary and primary reviewers corrected the same template segmentation, we felt that corrected segmentations were likely to inherit similarities from the template that would bias the test dataset toward the training dataset. Therefore, we chose to exclude the 86 primary reviewer segmentations that corresponded to cases with a secondary reviewer from the training data. The model was initially trained end-to-end by fivefold cross validation. This permitted fine-tuning of the hyperparameters (e.g., epochs, learning rate, batch size) without overfitting the external training set. The whole dataset split and secondary reviewer segmentations were recently made available through TCIA in an update (version 2) of the original dataset.



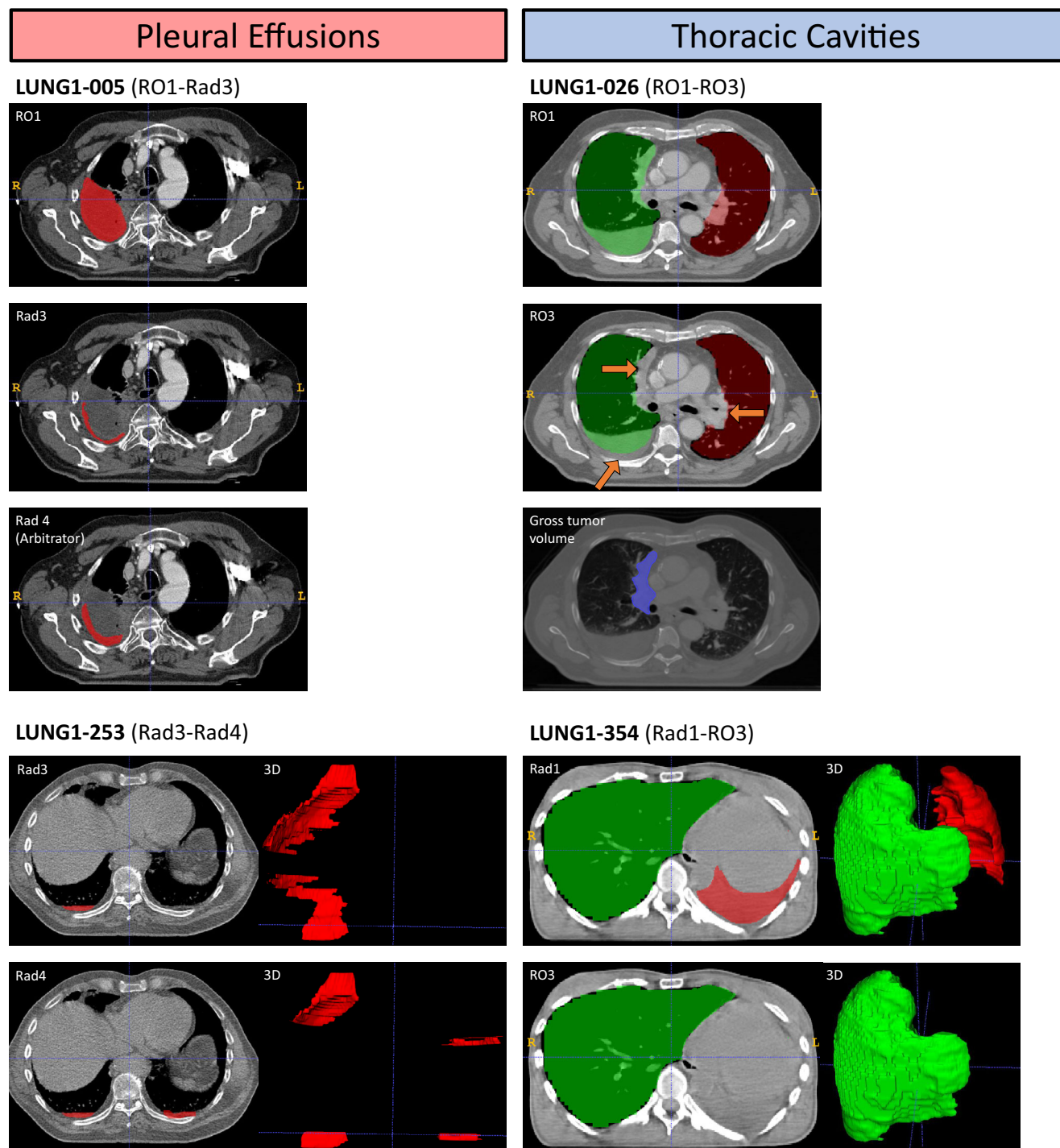


FIG. 6. A visual exploration of physician-corrected segmentation pairs with the least interobserver agreement. Case LUNG1-005 accounts for the worst Dice similarity coefficient (DSC), surface DSC, and  $\kappa$  values between any pair of pleural effusion segmentations. RO1 mistook atelectatic lung for effusion, but Rad3 and Rad4 did not. Case LUNG1-253 accounts for the worst 95HD and symmetric ASD values between any pair of pleural effusion segmentations. Rad3 and Rad4 varied in how much trace pleural fluid they chose to segment. This exposes a weakness in our pleural effusion segmentation methodology because we did not decide at projection initiation whether or to what extent trace pleural fluid should be part of the segmentation. Case LUNG1-026 accounts for the worst sDSC value between any pair of thoracic cavity segmentations. RO3 failed to segment the full extent of peri-mediastinal primary gross tumor volume, right effusion, and left hilum (orange arrows). Case LUNG1-354 accounts for the worst DSC,  $\kappa$ , 95HD, and symmetric ASD values between any pair of thoracic cavity segmentations. RO3 erroneously excluded a collapsed left lung. [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

To build the 3D U-Net model we used the architecture described by Çiçek et al.<sup>53</sup> For each epoch the train/validation split was 80% train (252 scans) and 20% validation (64

scans). The batch size was 1 scan, the DSC was used to evaluate loss, and the learning rate was set initially to 0.001 and adapted by the Adam optimization algorithm.<sup>57</sup> The DSC

plateaued at 100 epochs. The mean DSC performance on the test dataset was 0.95 (standard deviation  $\pm 0.05$ ) and 0.95 ( $\pm 0.04$ ) for the left and right lungs, respectively. For reference, the DSC between the primary and secondary reviewer segmentations was 0.98 ( $\pm 0.11$ ) and 0.98 ( $\pm 0.11$ ) for the left and right lungs, respectively.

To build the 2D U-Net model we used the same architecture as the 3D U-Net but changed all 3D convolution and deconvolution operations to 2D operations. In this case, the algorithm was trained on 2D axial slices (i.e., 256 by 256 voxels) rather than whole volumes. Like the 3D model, the 2D U-Net train/validation split was 80% train (32 256 slices) and 20% validation (8192 slices). The DSC plateaued at 60 epochs. The batch size was 64 slices, the DSC was used to evaluate loss, and the learning rate was set initially to 0.001 and adapted by the Adam optimization algorithm.<sup>57</sup> The mean DSC performance on the test dataset was 0.94 ( $\pm 0.10$ ) and 0.94 ( $\pm 0.10$ ) for the left and right lungs, respectively.

#### 4. DISCUSSION

To our knowledge, PleThora is the first public dataset of VOIs curated to capture all thoracic cavity pathologic change in patients with lung cancer, and the first public dataset of pleural effusion segmentations. The thoracic cavity segmentations are likely to be valuable to scientists and engineers who develop chest CT image processing pipelines that require robust methods to identify thoracic VOIs. We anticipate these segmentations to be a particularly useful addition to the corpus of training data for image processing pipelines, including the ones leveraging deep learning algorithms. Indeed, this project began as an effort to provide VOIs to study image feature symmetry between left and right thorax anatomy as a clinical outcomes predictor, building on previous work from our group that localized stroke cores by comparing and contrasting brain hemisphere information extracted by “symmetry-sensitive convolutional neural networks.”<sup>58</sup>

Our pleural effusion segmentations are likely to be useful for investigating two questions surrounding a CT or PET/CT finding of pleural effusion: (a) the prognostic significance of pleural effusion in various cancer types,<sup>59</sup> and (b) the capacity of CT to discriminate between benign and malignant effusions.<sup>60–63</sup> Regarding the first question, Ryu et al.<sup>59</sup> showed that in small cell lung cancer with stage I–III disease, the presence of even minimal pleural effusion confers an increased risk of death. Investigating the prognostic relationship of pleural effusion presence in other cancer types would presumably be facilitated by deep learning pleural effusion segmentation algorithms, such as might be trained with datasets like ours. Regarding the second question, some CT findings (nodular, mediastinal, parietal, and circumferential pleural thickening) are classically considered to be reasonably specific but poorly sensitive discriminators of malignant effusion.<sup>64</sup> Perhaps investigations into the value of pleural effusion quantitative imaging biomarkers for accomplishing this task could be fruitful. For example, Yang et al.<sup>65</sup> reported that

radiomics features extracted from lungs and pleura contained information capable of discriminating between patients with and without NSCLC dry pleural dissemination (AUC: 0.93; 95% CI: 0.891–0.968), which is a contraindication to primary tumor surgical resection that cannot always be detected by gross visualization. Our pleural effusion segmentations could enable other quantitative imaging biomarker studies. Importantly, our pleural effusion segmentations exclude primary tumor, as outlined in “RTSTRUCT” files, but users should not mistake this to mean that the effusions are necessarily benign. Microscopic tumor and macroscopic tumor below the threshold of radiologic detection are likely to exist in the effusions as delineated. It is also important to note that we failed to establish a guiding threshold for inclusion or exclusion of trace pleural fluid, an inconsistency that is reflected in several spatial similarity metrics we used to gauge interobserver variability. This stated, the effusion segmentations are still likely to be useful and may serve such research initiatives as correlating effusion parameters with clinical data available at the NSCLC-Radiomics “Radiomics Lung1.clinical-version3-Oct 2019.csv” spreadsheet or investigating differences in radiomics features between effusion and primary tumor.

We acknowledge the limiting inconsistencies of human-delineated segmentations, even those from trained radiologists or radiation oncologists. We also acknowledge intrinsic limitations in the metrics themselves. The DSC and  $\kappa$  are both artificially increased in large volumes, and the power of  $\kappa$  to penalize chance agreement is artificially decreased by the high number of true negatives in our segmentations (i.e., the high number of voxels that neither reviewer segmented as part of the target). We attempted to buffer this limitation by calculating and comparing sDSC values, which are less sensitive to variation in size. The 95 HD and symmetric ASD are not inflated by volume but are only snapshots of segmentations at their average (ASD) and near their worst (95HD) differences. Notwithstanding these limitations, we consider that the measures of interobserver variability obtained between radiologist and radiation oncologist reviewers justify acceptability of these segmentations for public use.

#### 5. CONCLUSIONS

We describe PleThora, a dataset of 402 expert-vetted thoracic cavity segmentations, 78 expert-vetted pleural effusion segmentations, and corresponding clinical and technical metadata made available to the public through TCIA at <https://doi.org/10.7937/tcia.2020.6c7y-gq39>.<sup>51</sup> These segmentations have value for preprocessing steps in image analysis pipelines built for fundamental quantitative imaging tasks, including but not limited to pathologic lung segmentation, lesion detection, and radiomics feature extraction.

#### ACKNOWLEDGMENTS

SMS is funded by a grant from the Swiss Cancer League (BIL KLS-4300-08-2017). CDF has received funding and

salary support unrelated to this project from: the National Institutes of Health (NIH) National Institute for Dental and Craniofacial Research Establishing Outcome Measures Award (1R01DE025248/R56DE025248) and an Academic Industrial Partnership Grant (R01DE028290); National Cancer Institute (NCI) Early Phase Clinical Trials in Imaging and Image-Guided Interventions Program (1R01CA218148); an NIH/NCI Cancer Center Support Grant (CCSG) Pilot Research Program Award from the UT MD Anderson CCSG Radiation Oncology and Cancer Imaging Program (P30CA016672) and an NIH/NCI Head and Neck Specialized Programs of Research Excellence (SPORE) Developmental Research Program Award (P50CA097007); National Science Foundation (NSF), Division of Mathematical Sciences, Joint NIH/NSF Initiative on Quantitative Approaches to Biomedical Big Data (QuBB) Grant (NSF 1557679); NSF Division of Civil, Mechanical, and Manufacturing Innovation (CMMI) standard grant (NSF 1933369) a National Institute of Biomedical Imaging and Bioengineering (NIBIB) Research Education Programs for Residents and Clinical Fellows Grant (R25EB025787-01); the NIH Big Data to Knowledge (BD2K) Program of the NCI Early Stage Development of Technologies in Biomedical Computing, Informatics, and Big Data Science Award (1R01CA214825). CDF has also received direct industry grant support, honoraria, and travel funding from Elekta AB. WJZ and LG are supported in part by the NIH (UL1TR003167) and the Cancer Prevention and Research Institute of Texas (RP 170668). WJZ is also supported by an NIH grant (R01AG066749), and LG is also supported by a Learning Healthcare Award funded by the UTHealth Center for Clinical and Translational Science (CCTS).

## CONFLICT OF INTEREST

The authors declare no conflict of interest relevant to this publication or the data therein described.

<sup>a)</sup>Authors to whom correspondence should be addressed. Electronic mails: Kendall.j.kiser@uth.tmc.edu, cdfuller@mdanderson.org, Luca.Giancardo@uth.tmc.edu; Telephones: 713.500.6500, 713.563.2334, 713.500.3609.

## REFERENCES

- Kumar V, Gu Y, Basu S, et al. Radiomics: the process and the challenges. *Magn Reson Imaging*. 2012;30:1234–1248.
- Zhang G, Jiang S, Yang Z, et al. Automatic nodule detection for lung cancer in CT images: a review. *Comput Biol Med*. 2018;103:287–300.
- Ardila D, Kiraly AP, Bharadwaj S, et al. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nat Med*. 2019;25:954–961.
- Hu S, Hoffman EA, Reinhardt JM. Automatic lung segmentation for accurate quantitation of volumetric X-ray CT images. *IEEE Trans Med Imaging*. 2001;20:490–498.
- Tong Y, Udupa JK, Odhner D, Wu C, Schuster SJ, Torigian DA. Disease quantification on PET/CT images without explicit object delineation. *Med Image Anal*. 2019;51:169–183.
- Mansoor A, Bagci U, Foster B, et al. Segmentation and image analysis of abnormal lungs at CT: current approaches, challenges, and future trends. *Radiographics*. 2015;35:1056–1076.
- Mansoor A, Bagci U, Xu Z, et al. A generic approach to pathological lung segmentation. *IEEE Trans Med Imaging*. 2014;33:2293–2310.
- Gordaliza PM, Munoz-Barrutia A, Abella M, Desco M, Sharpe S, Vaquero JJ. Unsupervised CT lung image segmentation of a mycobacterium tuberculosis infection model. *Sci Rep*. 2018;8:9802.
- Wang X, Teng P, Lo P, et al. High throughput lung and lobar segmentation by 2D and 3D CNN on chest CT with diffuse lung disease. In: Stoyanov D, Taylor Z, Kainz B, eds. *Image Analysis for Moving Organ, Breast, and Thoracic Images*. Cham: Springer; 2018. [https://doi.org/10.1007/978-3-030-00946-5\\_21](https://doi.org/10.1007/978-3-030-00946-5_21)
- Harrison AP, Xu Z, George K, Lu L, Summers RM, Mollura DJ. Progressive and multi-path holistically nested neural networks for pathological lung segmentation from CT images. In: Descoteaux M, Maier-Hein L, Franz A, Jannin P, Collins DL, Duchesne S, eds. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2017*. Cham: Springer; 2017. [https://doi.org/10.1007/978-3-319-66179-7\\_71](https://doi.org/10.1007/978-3-319-66179-7_71)
- Deen KJ, Ganesan R, Merline A. Fuzzy-C-means clustering based segmentation and CNN-classification for accurate segmentation of lung nodules. *Asian Pac J Cancer Prev*. 2017;18:1869–1874.
- Gudmundsson E, Straus CM, Armato SG 3rd. Deep convolutional neural networks for the automated segmentation of malignant pleural mesothelioma on computed tomography scans. *J Med Imaging (Bellingham)*. 2018;5:034503.
- Park B, Park H, Lee SM, Seo JB, Kim N. Lung segmentation on HRCT and volumetric CT for diffuse interstitial lung disease using deep convolutional neural networks. *J Digit Imaging*. 2019;32:1019–1026.
- Chen G, Xiang D, Zhang B, et al. Automatic pathological lung segmentation in low-dose CT image using Eigenspace sparse shape composition. *IEEE Trans Med Imaging*. 2019;38:1736–1749.
- Soliman A, Khalifa F, Elnakib A, et al. Accurate lungs segmentation on CT chest images by adaptive appearance-guided shape modeling. *IEEE Trans Med Imaging*. 2017;36:263–276.
- Doel T, Gavaghan DJ, Grau V. Review of automatic pulmonary lobe segmentation methods from CT. *Comput Med Imaging Graph*. 2015;40:13–29.
- Ren H, Zhou L, Liu G, et al. An unsupervised semi-automated pulmonary nodule segmentation method based on enhanced region growing. *Quant Imaging Med Surg*. 2020;10:233–242.
- Dong J, Lu K, Dai S, Xue J, Zhai R. Auto-Segmentation of Pathological Lung Parenchyma Based on Region Growing Method. Paper presented at: Internet Multimedia Computing and Service; Aug. 23–25, 2017; Qingdao, China. [https://doi.org/10.1007/978-981-10-8530-7\\_23](https://doi.org/10.1007/978-981-10-8530-7_23)
- Hofmanninger J, Prayer F, Pan J, Rohrich S, Prosch H, Langs G. Automatic lung segmentation in routine imaging is a data diversity problem, not a methodology problem. arXiv:2001.11767. Published 31 Jan 2020. Accessed 4 Apr 2020.
- Shaukat F, Raja G, Frangi AF. Computer-aided detection of lung nodules: a review. *J Med Imaging*. 2019;6:1.
- Clark K, Vendt B, Smith K, et al. The cancer imaging archive (TCIA): maintaining and operating a public information repository. *J Digit Imaging*. 2013;26:1045–1057.
- Aerts HJWL, Wee L, Rios Velazquez E, et al. Data from NSCLC-Radiomics [Dataset]. In: *The Cancer Imaging Archive*; 2019. <https://doi.org/10.7937/K9/TCIA.2015.PF0M9REJ>
- Aerts HJ, Velazquez ER, Leijenaar RT, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun*. 2014;5:4006.
- Li X, Morgan PS, Ashburner J, Smith J, Rorden C. dcm2niix.exe [computer program]. Version v1.0.201811142020. Accessed January 23, 2019. Available from: <https://www.nitrc.org/plugins/mwiki/index.php/dcm2ni:MainPage>
- Li X, Morgan PS, Ashburner J, Smith J, Rorden C. The first step for neuroimaging data analysis: DICOM to Nifti conversion. *J Neurosci Methods*. 2016;264:47–56.
- Pesuk V. Lung Segmentation (3D) [computer program]. GitHub 2017. Accessed December 15, 2018. Available from <https://github.com/imlab-uitp/lung-segmentation-3d>



27. Yushkevich PA, Piven J, Hazlett HC, et al. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *Neuroimage*. 2006;31:1116–1128.
28. Panchal A. Dicompyler [computer program]. Version vol 0.4.22014. Accessed February 4, 2020. Available from <https://www.dicompyler.com/>
29. Oliphant TE. Python for scientific computing. *Comput Sci Eng*. 2007;9:10–20.
30. Abraham A, Pedregosa F, Eickenberg M, et al. Machine learning for neuroimaging with scikit-learn. *Front Neuroinform*. 2014;8:14.
31. McKinney W. Data structures for statistical computing in Python. *Proceedings of the 9th Python in Science Conference (SCIPY 2010)*. 2010;445:51–56.
32. Nikolov S, Blackwell S, Mendes R, et al. Deep learning to achieve clinically applicable segmentation of head and neck anatomy for radiotherapy. arXiv:1809.04430. Published 12 Sep 2018. Accessed 07 Mar 2019.
33. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python. *J Mach Learn Res*. 2011;12:2825–2830.
34. Hunter JD. Matplotlib: a 2d graphics environment. *Comput Sci Eng*. 2007;9:90–95.
35. Waskom M, Botvinnik O, O’Kane D, et al. *MWaskom/Seaborn* [computer program]. Version 0.9.0: Zenodo; 2018. Accessed 03 Apr 2020. Available from <https://doi.org/10.5281/zenodo.1313201>
36. Dice LR. Measures of the amount of ecologic association between species. *Ecology*. 1945;26:297.
37. Cohen J. A coefficient of agreement for nominal scales. *Educational Psychol Meas*. 1960;20:37–46.
38. Huttenlocher DP, Klanderman GA, Rucklidge WJ. Comparing images using the Hausdorff distance. *IEEE Trans Pattern Anal Mach Intell*. 1993;15:850–863.
39. Yao J, Bliton J, Summers RM. Automatic segmentation and measurement of pleural effusions on CT. *IEEE Trans Bio-med Eng*. 2013;60:1834–1840.
40. von Falck C, Meier S, Jördens S, King B, Galanski M, Shin HO. Semi-automated segmentation of pleural effusions in MDCT datasets. *Acad Radiol*. 2010;17:841–848.
41. Anderson SA, Danelson KA, Mammarappallil JG, Chiles C, Stitzel JD. Semiautomatic method of quantifying pleural effusions using computed tomography scans - biomed 2013. *Biomed Sci Instrum*. 2013;49:13–19.
42. Song L, Gao J, Wang S, Hu H, Guo Y. A quantitative evaluation of pleural effusion on computed tomography scans using B-spline and local clustering level set. *J Xray Sci Technol*. 2017;25:887–905.
43. Moy MP, Levsky JM, Berko NS, Godelman A, Jain VR, Haramati LB. A new, simple method for estimating pleural effusion size on CT scans. *Chest*. 2013;143:1054–1059.
44. Brock KK, Mutic S, McNutt TR, Li H, Kessler ML. Use of image registration and fusion algorithms and techniques in radiotherapy: report of the AAPM Radiation Therapy Committee Task Group No. 132. *Med Phys*. 2017;44:e43–e76.
45. Yang J, Veeraraghavan H, Armato SG 3rd, et al. Autosegmentation for thoracic radiation treatment planning: a grand challenge at AAPM 2017. *Med Phys*. 2018;45:4568–4581.
46. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159.
47. Vaassen F, Hazelaar C, Vaniqui A, et al. Evaluation of measures for assessing time-saving of automatic organ-at-risk segmentation in radiotherapy. *Phys Imaging Radiat Oncol*. 2020;13:1–6.
48. Mann HB, Whitney DR. On a test of whether one of two random variables is stochastically larger than the other. *Ann Math Stat*. 1947;18:50–60.
49. Kiser K, Barman A, Stieb S, Fuller CD, Giancardo L. Novel autosegmentation spatial similarity metrics capture the time required to correct segmentations better than traditional metrics in a thoracic cavity segmentation workflow. medRxiv: <https://doi.org/10.1101/2020.05.14.20102103>. Published 18 May 2020. Accessed 18 May 2020.
50. Wilkinson MD, Dumontier M, Aalbersberg IJ, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data*. 2016;3:160018.
51. Kiser KJ, Ahmed S, Stieb SM, et al. Data from the thoracic volume and pleural effusion segmentations in diseased lungs for benchmarking chest CT processing pipelines [Dataset]. In: *The Cancer Imaging Archive*; 2020. <https://doi.org/10.7937/tcia.2020.6e7y-gq39>
52. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. Paper presented at: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015; 5–9 October, 2015; Munich, Germany. [https://doi.org/10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28)
53. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. Paper presented at: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016; 17–21 October, 2016; Athens, Greece. [https://doi.org/10.1007/978-3-319-46723-8\\_49](https://doi.org/10.1007/978-3-319-46723-8_49)
54. Milletari F, Navab N, Ahmadi S. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. Paper presented at: 2016 Fourth International Conference on 3D Vision (3DV); 25–28 Oct. 2016, 2016. <https://doi.org/10.1109/3DV.2016.79>
55. Zhou Z, Rahman Siddiquee MM, Tajbakhsh N, Liang J. UNet++: A Nested U-Net Architecture for Medical Image Segmentation. Paper presented at: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support; 2018//, 2018; Cham. [https://doi.org/10.1007/978-3-030-00889-5\\_1](https://doi.org/10.1007/978-3-030-00889-5_1)
56. Brosch T, Tang LYW, Yoo Y, Li DKB, Traboulsee A, Tam R. Deep 3D convolutional encoder networks with shortcuts for multiscale feature integration applied to multiple sclerosis lesion segmentation. *IEEE Trans Med Imaging*. 2016;35:1229–1239.
57. Kingma DP, Ba JL. Adam: A Method for Stochastic Optimization. Paper presented at: 3rd International Conference on Learning Representations, ICLR 2015; 7–9 May, 2015; San Diego, CA. arXiv:1412.6980v9.
58. Barman A, Inam ME, Lee S, Savitz S, Sheth S, Giancardo L. Determining ischemic stroke from CT-angiography imaging using symmetry-sensitive convolutional networks. 2019 IEEE 16th International Symposium on Biomedical Imaging; 2019. <https://doi.org/10.1109/ISBI.2019.8759475>
59. Ryu JS, Lim JH, Lee JM, et al. Minimal pleural effusion in small cell lung cancer: proportion, mechanisms, and prognostic effect. *Radiology*. 2016;278:593–600.
60. Zhang X, Duan H, Yu Y, et al. Differential diagnosis between benign and malignant pleural effusion with dual-energy spectral CT. *PLoS One*. 2018;13:e0193714.
61. Otsu T, Kataoka Y, Ikegaki S, et al. Pleural effusion biomarkers and computed tomography findings in diagnosing malignant pleural mesothelioma: a retrospective study in a single center. *PLoS One*. 2017;12:e0185850.
62. Herrera Lara S, Fernandez-Fabrellas E, Juan Samper G, et al. Predicting malignant and paramalignant pleural effusions by combining clinical, radiological and pleural fluid analytical parameters. *Lung*. 2017;195:653–660.
63. Nakajima R, Abe K, Sakai S. Diagnostic ability of FDG-PET/CT in the detection of malignant pleural effusion. *Medicine (Baltimore)*. 2015;94:e1010.
64. Hallifax RJ, Talwar A, Wrightson JM, Edey A, Gleeson FV. State-of-the-art: radiological investigation of pleural disease. *Respir Med*. 2017;124:88–99.
65. Yang M, Ren Y, She Y, et al. Imaging phenotype using radiomics to predict dry pleural dissemination in non-small cell lung cancer. *Ann Transl Med*. 2019;7:259.

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**Appendix S1.** Data dictionary to the “Thorax and Pleural Effusion Segmentation Metadata” spreadsheet.