

CT images with expert manual contours of thoracic cancer for benchmarking auto-segmentation accuracy

Citation for published version (APA):

Yang, J., Veeraraghavan, H., van Elmpt, W., Dekker, A., Gooding, M., & Sharp, G. (2020). CT images with expert manual contours of thoracic cancer for benchmarking auto-segmentation accuracy. *Medical Physics*, 47(7), 3250-3255. <https://doi.org/10.1002/mp.14107>

Document status and date:

Published: 01/07/2020

DOI:

[10.1002/mp.14107](https://doi.org/10.1002/mp.14107)

Document Version:

Publisher's PDF, also known as Version of record

Document license:

Taverne

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

CT images with expert manual contours of thoracic cancer for benchmarking auto-segmentation accuracy

Jinzhong Yang^{a)}

Department of Radiation Physics, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

Harini Veeraraghavan

Department of Medical Physics, Memorial Sloan Kettering Cancer Centre, New York, NY, USA

Wouter van Elmpt and Andre Dekker

Department of Radiation Oncology (MAASTRO), GROW - School for Oncology and Developmental Biology, Maastricht University Medical Center, Maastricht, The Netherlands

Mark Gooding

Mirada Medical Ltd, Oxford, UK

Greg Sharp

Department of Radiation Oncology, Massachusetts General Hospital, Boston, MA, USA

(Received 20 December 2019; revised 17 February 2020; accepted for publication 22 February 2020; published 28 March 2020)

Purpose: Automatic segmentation offers many benefits for radiotherapy treatment planning; however, the lack of publicly available benchmark datasets limits the clinical use of automatic segmentation. In this work, we present a well-curated computed tomography (CT) dataset of high-quality manually drawn contours from patients with thoracic cancer that can be used to evaluate the accuracy of thoracic normal tissue auto-segmentation systems.

Acquisition and validation methods: Computed tomography scans of 60 patients undergoing treatment simulation for thoracic radiotherapy were acquired from three institutions: MD Anderson Cancer Center, Memorial Sloan Kettering Cancer Center, and the MAASTRO clinic. Each institution provided CT scans from 20 patients, including mean intensity projection four-dimensional CT (4D CT), exhale phase (4D CT), or free-breathing CT scans depending on their clinical practice. All CT scans covered the entire thoracic region with a 50-cm field of view and slice spacing of 1, 2.5, or 3 mm. Manual contours of left/right lungs, esophagus, heart, and spinal cord were retrieved from the clinical treatment plans. These contours were checked for quality and edited if necessary to ensure adherence to RTOG 1106 contouring guidelines.

Data format and usage notes: The CT images and RTSTRUCT files are available in DICOM format. The regions of interest were named according to the nomenclature recommended by American Association of Physicists in Medicine Task Group 263 as Lung_L, Lung_R, Esophagus, Heart, and SpinalCord. This dataset is available on The Cancer Imaging Archive (funded by the National Cancer Institute) under Lung CT Segmentation Challenge 2017 (<http://doi.org/10.7937/K9/TCIA.2017.3r3fvz08>).

Potential applications: This dataset provides CT scans with well-delineated manually drawn contours from patients with thoracic cancer that can be used to evaluate auto-segmentation systems. Additional anatomies could be supplied in the future to enhance the existing library of contours. © 2020 American Association of Physicists in Medicine [<https://doi.org/10.1002/mp.14107>]

Key words: automatic segmentation, CT, grand challenge, radiation therapy, thoracic cancer

1. INTRODUCTION

Contouring on computed tomography (CT) scans is an important task in radiation treatment planning for achieving favorable dose distributions to targets with minimal radiation toxicity to adjacent normal tissues.¹ Manual delineation of targets and organs at risk is both time consuming and subject to interobserver variability.^{2,3} In the past decade, the development of computer-aided auto-segmentation has gradually changed the traditional way of contouring for treatment planning. Auto-segmentation has proven an effective for reducing uncertainties in contouring and improving the efficiency of the contouring process.^{4,5} Many auto-segmentation

algorithms have been developed and evaluated with CT images for treatment planning.⁶⁻¹² However, the use of auto-segmentation in the clinic is still limited, in part because auto-segmentation is less accurate or robust than manual segmentation in some cases, and because benchmark datasets are not publicly available for commissioning auto-segmentation tools.

In recent years, the concept of a “grand challenge” has emerged as an unbiased and effective approach for evaluating different segmentation approaches.^{8,9,13} An unbiased evaluation also relies on the quality of the curated benchmark dataset. Organizers of such challenges must take care in choosing the benchmark dataset and performing thorough quality checks on the reference contours. In this article, we present a

dataset originally prepared for the Thoracic Auto-Segmentation Challenge held at the 2017 Annual Meeting of the American Association of Physicists in Medicine (AAPM). Details of the preparations for this challenge, including quality assurance of manual contours and challenge results, are reported elsewhere⁹; this paper focuses on describing the image data acquisition and use of the presented data. We have made this dataset publicly available through The Cancer Imaging Archive (TCIA)¹⁴ for researchers in medical physics or medical image analysis to develop and evaluate their own segmentation methods.

This dataset consists of CT scans with manually drawn contours for use in testing thoracic normal tissue auto-segmentation algorithms. These scans were acquired for treatment planning purposes, with each patient in the treatment planning position with cradles used for immobilization. The image quality is sufficient for contouring both tumors and organs at risk for treatment planning. This dataset can help address the lack of publicly available datasets to facilitate the testing of commercial and in-house segmentation algorithms. We believe this dataset is a valuable resource to the medical physics and medical image analysis communities.

2. ACQUISITION AND VALIDATION METHODS

2.A. Overview of the dataset

The dataset consists of treatment simulation CT scans obtained before radiotherapy that include manually drawn contours from 60 patients with thoracic cancer; scans were obtained from three institutions: MD Anderson Cancer Center (MDACC), Memorial Sloan Kettering Cancer Center (MSKCC), and the MAASTRO clinic. Each institution provided scans from 20 patients, including mean intensity projection on four-dimensional CT (4D CT) scans, the exhale phase on 4D CT scans, and free-breathing CT scans depending on their clinical practice. All CT scans covered the entire thoracic region. Scans from patients with collapsed lungs or with the esophagus terminating superior to the lower lobes of lungs were excluded. The median patient age was 70 yr (range 37–92 yr) at the time of CT scanning; 33 patients (55%) were men and 27 (45%) were women. The purpose of this dataset was to provide benchmarks for thoracic normal tissue segmentations for the Thoracic Auto-Segmentation Challenge at the 2017 AAPM Annual Meeting (<http://autocontouringchallenge.org/>) (accessed on December 9, 2019).

2.A.1. MDACC datasets

Under an institutional review board—approved protocol and waiver of informed consent (RCR03-0400), we retrospectively retrieved CT scans, with manual contours, from 20 patients with nonsmall cell lung cancer (NSCLC) who received definitive chemoradiation at The University of Texas MD Anderson Cancer Center from August 2009 to December 2012. Patients were selected at random without specific

restrictions on sex, age, or body anatomy. The median patient age was 66.5 yr (range 47–80 yr) at the time of CT scanning; 13 patients (65%) were men and 7 (35%) were women. Setup for all patients took place while the patients were supine, with their arms above their heads, and immobilized with custom Vac-Lok cradles (Civco Medical Solutions, Kalona, IA).

All patients underwent 4D CT scanning on a GE Light-Speed 16-slice scanner (GE Medical Systems, Waukesha, WI) used for radiotherapy simulation at MD Anderson Cancer Center. All scans covered the entire thoracic region and were based on a 4D CT scanning protocol with 120 kVp, 50 mAs, a 50-cm field of view, and a 2.5-mm slice thickness. Images were reconstructed by using a standard convolution kernel to a matrix size of 512×512 . The number of slices ranged from 120 to 176, and slice spacing was 2.5 mm for all 20 patients. Ten phases of CT scans were reconstructed from each 4D CT scan set. The mean intensity projection of the 10 phase CT scans (i.e., the averaged 4D CT) was acquired and included in the dataset because it is used in our clinic for radiotherapy treatment planning. All clinical contours were drawn on the averaged 4D CT in a Pinnacle treatment planning system (Philips Medical Systems, Fitchburg, WI). Lung contours were drawn with fusion of the 30% phase CT, and other contours (esophagus, heart, and spinal cord) were drawn with fusion of the 50% phase CT (end-of-exhalation phase).

2.A.2. MSKCC datasets

CT scans were retrospectively retrieved from 20 patients with NSCLC treated with radiation therapy at Memorial Sloan Kettering Cancer Center after approval and waiver of informed consent by the institutional review board. The patients had been treated from January 2015 to December 2016 and were selected at random without specific restrictions on sex, age, or body anatomy. The median patient age was 73.5 yr (range 70.25–83.0 yr) at the time of CT scanning; 8 patients (40%) were men and 12 (60%) were women. CT scans were acquired while the patients were supine, with arms above their heads, during treatment simulation.

Nine of the 20 patients underwent 4D CT scanning on a Philips Brilliance Big Bore 16-slice scanner (Philips Medical Systems, Cleveland, OH) used for radiotherapy simulation at Memorial Sloan Kettering Cancer Center; 10 patients underwent fast helical free-breathing CT scans on a Philips Brilliance Big Bore 16 slice scanner; and the remaining patient underwent deep inspiration breath-hold CT scan with contrast on a GE Discovery ST (General Electric) scanner. All scans covered the entire thoracic region with 120 kVp, 50 mAs, a 65-cm field of view, and a 3-mm slice thickness. Images were reconstructed by using a standard convolution kernel to a matrix size of 512×512 . The number of slices ranged from 103 to 279, with an average slice spacing of 2.5 mm for all 20 patients. For patients who underwent 4D CT scanning, ten phases of CT scans were reconstructed, and the mean intensity projection of the ten CT phases (i.e., the averaged 4D CT) was acquired and included in the dataset. All clinical

contours were drawn on the averaged 4D CT scan or the free-breathing CT scan in a Varian Eclipse treatment planning system (Varian Medical Systems, Palo Alto, CA).

2.A.3. MAASTRO datasets

CT scans were retrospectively retrieved from 20 patients with NSCLC treated with radiation therapy at MAASTRO clinic (Maastricht, The Netherlands) after approval and waiver of informed consent by the institutional review board. The patients had been treated from July 2015 to January 2016 and were selected at random without specific restrictions on sex, age, or body anatomy. The median patient age was 68.5 yr (range 49–89 yr) at the time of CT scanning; 12 patients (60%) were men and 8(40%) were women. CT scans were acquired while the patients were supine with their arms above their heads during treatment simulation.

Patients underwent 4D CT scanning on one of two Siemens scanners (Siemens Biograph 40 PET/CT or a Siemens Sensation Open CT scanner; Siemens Healthineers, Forchheim, Germany) dedicated to radiotherapy simulation. All scans covered the entire thoracic region and involved a 4D CT scanning protocol with 120 kVp, variable mAs settings (mainly at 100 mAs [range 100–240 mAs]), a 50-cm field of view (which could be extended to 65 cm as needed), and a 3-mm slice thickness. Images were reconstructed by using a standard convolution kernel B19f or B30f to a matrix size of 512×512 and variable numbers of slices between patients depending on the height of the patient. Eight phases of CT scans were reconstructed for each 4D CT scan and the 50% exhale phase was used for delineation and radiotherapy treatment planning dose calculations. All clinical contours were drawn on the 50% Exhale CT phase (mid-ventilation) in a Varian Eclipse treatment planning system (Varian Medical Systems, Palo Alto, CA).

2.B. Manual contours

Manually drawn contours were collected from clinical treatment plans from all three institutions for this dataset. Contours of the following structures were included: left and right lungs, esophagus, heart, and spinal cord. The clinical contours were quality checked by one experienced clinical medical physicist (GS) and edited to adhere to RTOG 1106 contouring guidelines^{15,16} as closely as possible. The editing specifically did not remove interinstitutional variability in the original contours. The dataset is expected to include several areas of inconsistency. Detailed contouring notes are described elsewhere⁹ and are briefly summarized below.

- *Lungs*: Left lung and right lung were contoured into two regions of interest (ROIs). Tumor was excluded from the lung contours, but the extent of exclusion could differ in scans from different institutions. Hilar airways and vessels > 5 mm were excluded. The main bronchi were always excluded but secondary bronchi might be included in some patients, based on the

original clinical contours. Inclusion or exclusion of small vessels were not edited in general, also based on the original clinical contours.

- *Esophagus*: The superior end of esophagus was edited to be in the first slice of the cricoid cartilage, and the inferior end of esophagus was edited to be in the first slice of the gastroesophageal junction. Original clinical contours within the range were generally not edited.
- *Heart*: The superior border was edited for consistency with RTOG guidelines for all patients; however, inclusion or exclusion of the pericardial sac was inconsistent in some cases.
- *Spinal cord*: The original clinical contours of spinal cord might have extended beyond the cricoid superiorly and beyond the L2 inferiorly. These contours were not edited. For cases without enough contoured slices, the superior border was extended to the bottom slice of the cricoid and the inferior border was extended to the bottom slice of the L2.

These anatomic structures were selected because they are important organs at risk in radiotherapy treatment planning and are of particular interest in plan evaluation. The spinal cord is a critical organ that must be protected from excessive dose, as is the heart; radiation-induced cardiac toxicity and mortality is an important concern for dose escalation in lung cancer radiotherapy.¹⁷⁻¹⁹ Furthermore, higher dose to the lungs is known to be associated with higher risk of radiation-induced pneumonitis,²⁰ and higher dose to the esophagus with higher risk of radiation-induced esophagitis.²¹ Manual contours from three patients, one from each institution, are shown in Fig. 1. Volume statistics for these structures in all 60 patients are shown in Table I.

3. DATA FORMAT AND USAGE NOTES

This data collection is available at The Cancer Imaging Archive (TCIA; <https://www.cancerimagingarchive.net/>) in the collection **Lung CT Segmentation Challenge 2017** (<http://doi.org/10.7937/K9/TCIA.2017.3r3fvz08>).²² Both the CT and structure contour files are provided in DICOM, which is used by most commercial treatment planning systems. After performing a quality check of all manual contours and CT images, we used a customized version of the Clinical Trial Processor (<https://mirc.rsna.org>) provided by TCIA to anonymize and transfer the data to the TCIA servers. After that, TCIA curators performed additional review to ensure full anonymization in compliance with HIPAA regulations and to keep critical information necessary for analysis. A DICOM RT structure file was associated with each CT scan, and this file contained five ROIs: left/right lungs, esophagus, heart, and spinal cord. These ROIs are named as Lung_L, Lung_R, Esophagus, Heart, and SpinalCord according to the nomenclature recommended by AAPM Task Group 263.²³

Because this dataset was used for the AAPM Thoracic Auto-Segmentation Challenge,⁹ the dataset was separated into three data groups, stratified per institution, with 36

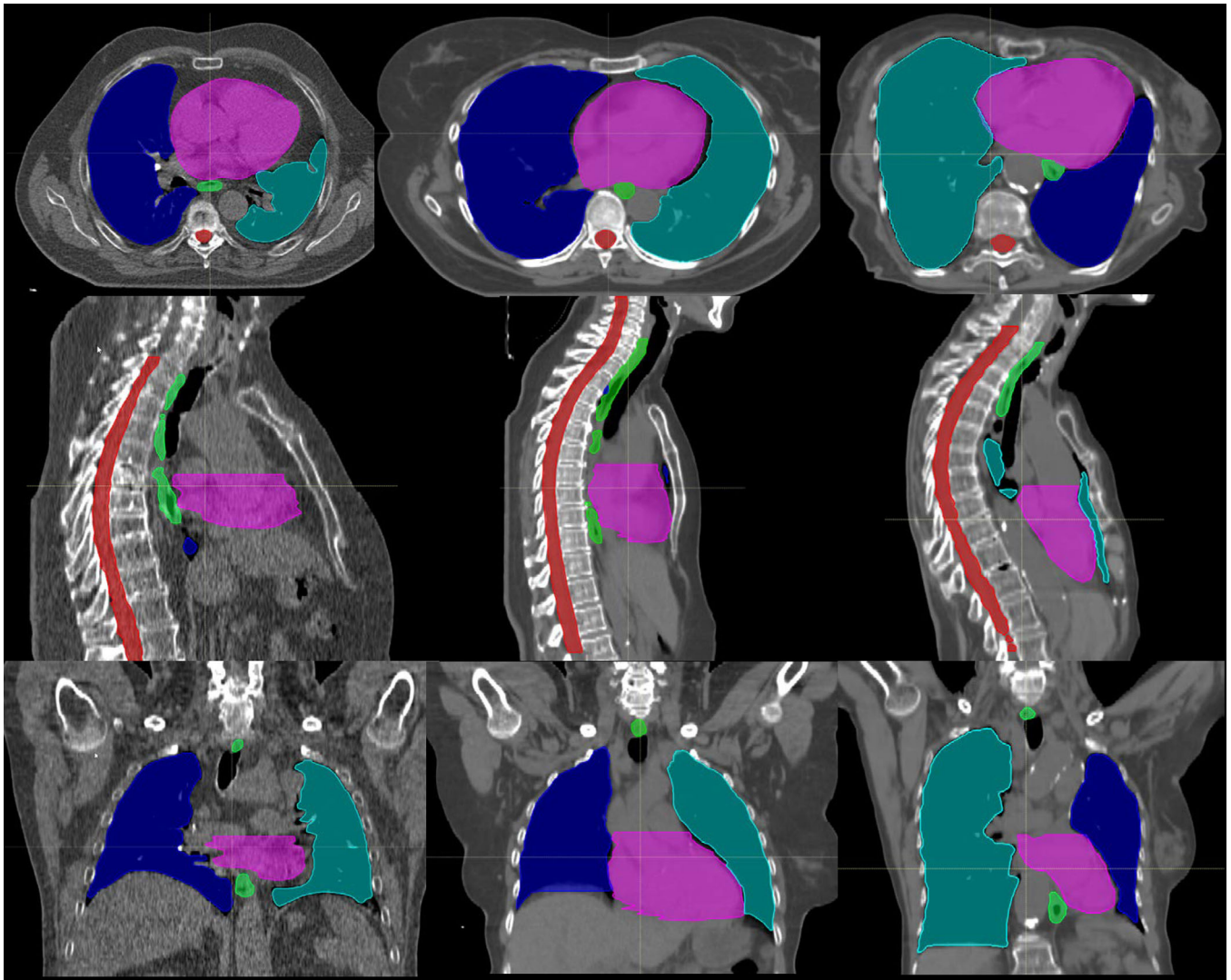


FIG. 1. Axial (top row), sagittal (middle row), and coronal (bottom row) views of contoured organs for three patients (left lung [light blue], right lung [dark blue], esophagus [green], heart [pink], spinal cord [red]), each from a different institution. [Color figure can be viewed at wileyonlinelibrary.com]

TABLE I. Volume statistics for the contoured structures on computed tomography scans from all 60 patients. SD, standard deviation.

	Minimum volume (cm ³)	Maximum volume (cm ³)	Median volume (cm ³)	Mean volume (cm ³)	Standard deviation
Left lung	128.1	4091.0	1564.1	1644.8	724.3
Right lung	892.4	3565.8	1910.2	2026.1	598.8
Esophagus	29.8	100.9	47.1	49.1	13.2
Heart	457.0	1352.2	741.0	772.5	196.4
Spinal cord	31.8	109.6	72.1	72.8	17.9

patients in a training dataset, 12 patients in an offsite test dataset, and 12 patients in a live test dataset. Patient ID numbers were used to distinguish each dataset and institution. For example, in the training dataset, patient IDs were coded as *LCTSC-Train-Sm-nnn*, where *Sm* (*m* = 1, 2, 3) stands for institution (*S1* = MAASTRO, *S2* = MDACC, *S3* = MSKCC)

and *nnn* (*nnn* = 001, 002, . . . , 012) stands for the dataset ID at one institution. Similarly, the patient IDs for the offsite test dataset were coded as *LCTSC-Test-Sm-10n*, and those for the live test data were *LCTSC-Test-Sm-20n*.

Data can be downloaded from the TCIA by using the NBIA Data Retriever.²⁴ Two directory types can be chosen from the interface of the NBIA Data Retriever. In the “Classic Directory Name,” the images are saved in the subfolder of the specified folder as *CollectionName\PatientID\StudyInstanceUID\SeriesInstanceUID\xxxxxx.dcm*, where *xxxxxx* starts from 000000 to the total number of image slices minus 1. Each image slice is stored in one DICOM file. By contrast, selection of the “Descriptive Directory Name,” *StudyInstanceUID* and *SeriesInstanceUID* in the folder name will be replaced by a combination of study date, study ID, study description, and *StudyInstanceUID*, and a combination of series number, series description, and *SeriesInstanceUID*, respectively.²⁴ The RTSTRUCT DICOM file can be found under the same structured format [*CollectionName\PatientID*

`\StudyInstanceUID\SeriesInstanceUID\000000.dcm`]; however, it should be noted that the SeriesInstanceUID is different than that of the DICOM images of the same patient.

4. DISCUSSION

Auto-segmentation has become increasingly useful in radiation oncology and is the subject of great interest in medical physics.^{6,9,11} As deep learning has become more popular, many researchers have devoted themselves to developing new and more accurate auto-segmentation algorithms for radiation oncology applications. Deep learning algorithms require a great amount of well-curated data for training if high-quality segmentation results are to be achieved. The dataset we introduce here represents the first attempt to supply a dedicated dataset for the evaluation of auto-segmentation algorithms for thoracic radiotherapy treatment planning. We expect this dataset to be useful, as it will provide well-curated contours for a relatively large group of patients with thoracic cancer.

As noted earlier, this dataset was originally prepared for the Thoracic Auto-Segmentation Challenge which took place at the 2017 AAPM Annual Meeting in Denver, CO.⁹ More than 100 participants registered for the challenge during the offline contest at the challenge website (<http://autocontouringchallenge.org/>) and used the training dataset to train their segmentation algorithms. Seven teams participated in the live contest in August 2017 in Denver. Of the seven teams, five developed deep learning auto-segmentation algorithms using the supplied training dataset. The top 3 segmentation results (all of which used deep learning) showed more consistent contouring than the measured interobserver variability, which supports the assumption that the number of patients was sufficiently large to capture anatomic variability among patients in training a good deep learning model for segmentation. Also, the interinstitutional variability in the dataset may well be helpful in training a deep learning model. Another report on the challenge results also shows how this dataset was used to train a deep learning model effectively for auto-segmentation.²⁵ However, we acknowledge that the size of the current dataset is limited for training and evaluating a large deep learning model.

This dataset has several potential future applications. First, this well-curated dataset is a good resource for researchers to develop innovative auto-segmentation algorithms. This dataset can be used together with their own data to train their algorithms, by increasing the data variability to improve their deep learning models. Next, since this dataset is collected from several institutions, it may be useful for studying interinstitutional contouring variability. An example of such a study would involve taking this dataset, contouring the tumor, and creating treatment plans for these patients, followed by evaluating the dosimetric impact of variability in contouring. Finally, this dataset could also serve as a benchmark for commissioning commercial auto-segmentation tools for thoracic radiotherapy treatment planning, provided that the commercial software had not used this dataset to train the algorithm.

This published dataset has some limitations. We did not remove all interinstitutional variations when we checked the quality of the clinical contours and thus some areas in the dataset could harbor inconsistencies. Moreover, the CT scans in this dataset were acquired for the purpose of radiotherapy simulation and treatment planning. Patients were immobilized with a specific device, thereby reducing the possible degrees of freedom in patient positioning. This limits the applications of this dataset and makes it more suitable for radiation oncology applications than others. Also, the in-slice resolution was about 1 mm, which may not be sufficient for radiology applications. In addition, tumors in some patients may have affected the shape and size of normal anatomic structures, thereby rendering the contours inconsistent.

5. CONCLUSIONS

We described a dataset of well-curated thoracic CT scans from 60 patients with thoracic cancer, with high-quality manually drawn contours of four important organs at risk in radiotherapy treatment planning. This dataset is made publicly available through The Cancer Imaging Archive in DICOM format. We believe this dataset will be helpful to the medical physics community in evaluating auto-segmentation algorithms for delineating thoracic normal tissues.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

The authors thank the American Association of Physicists in Medicine (AAPM) for supporting and sponsoring the 2017 AAPM Thoracic Auto-segmentation Challenge and The Cancer Imaging Archive (TCIA) (funded by the National Cancer Institute) for hosting the datasets and making them available to the public. We further thank Artem Mamonov and Andrew Beers from Harvard Medical School for providing valuable support for the challenge website, Kirk Smith and Tracy Nolan from the University of Arkansas for Medical Sciences for data curation to TCIA, Tim Lustberg from Maastricht Clinic for data collection, and Christine Wogan from MD Anderson Cancer Center for reviewing the manuscript. Dr. Yang is supported in part by the National Institutes of Health Cancer Center Support (Core) Grant no. P30 CA016672 to The University of Texas MD Anderson Cancer Center. Dr. Veeraghavan is supported in part by the National Institutes of Health Cancer Center Support (Core) Grant P30 CA008748 (to the institution). Dr. Dekker acknowledges support from National Institutes of Health grant no. U01 CA143062 (Radiomics of NSCLC).

^{a)}Author to whom correspondence should be addressed. Electronic mail: jjyang4@mdanderson.org.

REFERENCES

1. Ezzell GA, Galvin JM, Low D, et al. Guidance document on delivery, treatment planning, and clinical implementation of IMRT: report of the IMRT subcommittee of the AAPM radiation therapy committee. *Med Phys.* 2003;30:2089–2115.
2. Stapleford LJ, Lawson JD, Perkins C, et al. Evaluation of automatic atlas-based lymph node segmentation for head-and-neck cancer. *Int J Radiat Oncol Biol Phys.* 2010;77:959–966.
3. Yang J, Wei C, Zhang L, Zhang Y, Blum RS, Dong L. A statistical modeling approach for evaluating auto-segmentation methods for image-guided radiotherapy. *Comput Med Imaging Graph.* 2012;36:492–500.
4. Teguh DN, Levendag PC, Voet PW, et al. Clinical validation of atlas-based auto-segmentation of multiple target volumes and normal tissue (swallowing/mastication) structures in the head and neck. *Int J Radiat Oncol Biol Phys.* 2011;81:950–957.
5. McCarroll R, Beadle B, Balter P, et al. Retrospective validation and clinical implementation of automated contouring of organs at risk in the head and neck: a step toward automated radiation treatment planning for low- and middle-income countries. *J Global Oncol.* 2018;4:1–11.
6. Sharp G, Fritscher KD, Pekar V, et al. Vision 20/20: perspectives on automated image segmentation for radiotherapy. *Med Phys.* 2014;41:050902.
7. Zhou R, Liao Z, Pan T, et al. Cardiac atlas development and validation for automatic segmentation of cardiac substructures. *Radiother Oncol.* 2017;122:66–71.
8. Raudaschl PF, Zaffino P, Sharp GC, et al. Evaluation of segmentation methods on head and neck CT: auto-segmentation challenge 2015. *Med Phys.* 2017;44:2020–2036.
9. 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.
10. Yang J, Haas B, Fang R, et al. Atlas ranking and selection for automatic segmentation of the esophagus from CT scans. *Phys Med Biol.* 2017;62:9140–9158.
11. Cardenas CE, Yang J, Anderson BM, Court LE, Brock KB. Advances in auto-segmentation. *Semin Radiat Oncol.* 2019;29:185–197.
12. Rhee D, Cardenas C, Elhalawani H, et al. Automatic detection of contouring errors using convolutional neural networks. *Med Phys.* 2019;46:E282–E282.
13. Pekar V, Allaire S, Qazi AA, Kim JJ, Jaffray DA. Head and neck auto-segmentation challenge: segmentation of the parotid glands. In van Ginneken B, Murphy K, Heimann T, Pekar V, Deng X, eds. *Medical Image Analysis for the Clinic: A Grand Challenge*; 2010:273–280. Beijing, China.
14. 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.
15. Atlases for organs at risk (OARs) in thoracic radiation therapy. <https://www.rtog.org/LinkClick.aspx?fileticket=qlz0qMZXfQs%3d&tabid=361>. Accessed: December 09, 2019.
16. Kong FM, Ritter T, Quint DJ, et al. Consideration of dose limits for organs at risk of thoracic radiotherapy: atlas for lung, proximal bronchial tree, esophagus, spinal cord, ribs, and brachial plexus. *Int J Radiat Oncol Biol Phys.* 2011;81:1442–1457.
17. Hardy D, Liu CC, Cormier JN, Xia R, Du XL. Cardiac toxicity in association with chemotherapy and radiation therapy in a large cohort of older patients with non-small-cell lung cancer. *Ann Oncol.* 2010;21:1825–1833.
18. Bradley JD, Paulus R, Komaki R, et al. Standard-dose versus high-dose conformal radiotherapy with concurrent and consolidation carboplatin plus paclitaxel with or without cetuximab for patients with stage IIIA or IIIB non-small-cell lung cancer (RTOG 0617): a randomised, two-by-two factorial phase 3 study. *Lancet Oncol.* 2015;16:187–199.
19. Luo Y, Xu Y, Liao Z, et al. Automatic segmentation of cardiac substructures from noncontrast CT images: accurate enough for dosimetric analysis? *Acta Oncol.* 2019;58:81–87.
20. Jin HK, Tucker SL, Liu HH, et al. Dose-volume thresholds and smoking status for the risk of treatment-related pneumonitis in inoperable non-small cell lung cancer treated with definitive radiotherapy. *Radiother Oncol.* 2009;91:427–432.
21. Niedzielski JS, Yang J, Stingo F, et al. Objectively quantifying radiation esophagitis with novel computed tomography-based metrics. *Int J Radiat Oncol Biol Phys.* 2016;94:385–393.
22. Yang J, Sharp G, Veeraraghavan H, et al. Data from lung CT segmentation challenge. The Cancer Imaging Archive. 2017. <http://doi.org/10.7937/K9/TCIA.2017.3r3fvz08>.
23. Mayo CS, Moran JM, Bosch W, et al. American association of physicists in medicine Task Group 263: standardizing nomenclatures in radiation oncology. *Int J Radiat Oncol Biol Phys.* 2018;100:1057–1066.
24. Downloading TCIA Images, <https://wiki.cancerimagingarchive.net/display/NBIA/TCIA+Current+Help+Topics#TCIACurrentHelpTopics-DownloadingTCIAImages>. Accessed: January 27, 2020.
25. Feng X, Qing K, Tustison NJ, Meyer CH, Chen Q. Deep convolutional neural network for segmentation of thoracic organs-at-risk using cropped 3D images. *Med Phys.* 2019;46:2169–2180.