Deep Learning–based Automatic Lung Segmentation on Multiresolution CT Scans from Healthy and Fibrotic Lungs in Mice

Francesco Sforazzini, MSc • Patrick Salome, MSc • Mahmoud Moustafa, PhD • Cheng Zhou, MD, PhD • Christian Schweiger, PhD • Katrin Rein, PhD • Nina Bougatz, PhD • Andreas Kudak, MSc • Henry Woodruff, PhD • Ludwig Dubois, PhD • Philippe Lambin, MD, PhD • Jürgen Debus, MD, PhD • Amir Abdollahi, MD, PhD • Maximilian Knoll, MD, PhD

Radiology: Artificial Intelligence 2022; 4(2):e210095 • https://doi.org/10.1148/ryai.210095 • Content codes: [AI] [CH] [CT]

Purpose: To develop a model to accurately segment mouse lungs with varying levels of fibrosis and investigate its applicability to mouse images with different resolutions.

Materials and Methods: In this experimental retrospective study, a U-Net was trained to automatically segment lungs on mouse CT images. The model was trained (n = 1200), validated (n = 300), and tested (n = 154) on longitudinally acquired and semiautomatically segmented CT images, which included both healthy and irradiated mice (group A). A second independent group of 237 mice (group B) was used for external testing. The Dice score coefficient (DSC) and Hausdorff distance (HD) were used as metrics to quantify segmentation accuracy. Transfer learning was applied to adapt the model to high-spatial-resolution mouse micro-CT segmentation (n = 20; group C [n = 16 for training and n = 4 for testing]).

Results: The trained model yielded a high median DSC in both test datasets: 0.984 (interquartile range [IQR], 0.977–0.988) in group A and 0.966 (IQR, 0.955–0.972) in group B. The median HD in both test datasets was 0.47 mm (IQR, 0.30–0.32 mm [group A]) and 0.47 mm (IQR, 0.30–0.32 mm [group B]). Spatially resolved quantification of differences toward reference masks revealed two hot spots close to the air-tissue interfaces, which are particularly prone to deviation. Finally, for the higher-resolution mouse CT images, the median DSC was 0.905 (IQR, 0.902–0.929) and the median 95th percentile of the HD was 0.33 mm (IQR, 0.261–2.78 mm).

Conclusion: The developed deep learning–based method for mouse lung segmentation performed well independently of disease state (healthy, fibrotic, emphysematous lungs) and CT resolution.

Supplemental material is available for this article.

Published under a CC BY 4.0 license.

Radiation therapy is an integral part of cancer treatment and aims for precise, high-dose irradiation of tumors to achieve a high probability of tumor control. Normal tissue must be spared to avoid tissue toxicity and to lower the rate of potential complications. For treatments within the thoracic region, it is important to account for the fact that the lung, together with the heart and spinal cord, is a dose-limiting organ. The main relevant adverse effects of irradiation are pneumonitis (short-term) and lung fibrosis (long-term).

Rodent animal models are established for the preclinical study of pathophysiologic processes involved in the development of lung fibrosis, which can be monitored with clinical CT or micro-CT. Lung volume or specific metrics, such as the fibrosis index, can be used as a surrogate for lung fibrosis (1). However, such analyses rely on accurate lung segmentation, irrespective of the degree of fibrosis or otherwise altered lung parenchyma (eg, emphysema).

Segmentation can be performed by different methods, with manual segmentation having the disadvantages of being time-consuming and lacking consensus guidelines for segmentation (eg, inclusion and exclusion of the bronchial system); such variability may impair interpretability and reproducibility. Semiautomatic approaches use Hounsfield unit thresholds to automatically identify low-attenuation lung parenchyma followed by manual curation (1,2). Finally, atlas-based registration (3,4) and multiorgan segmentation following injection of contrast agent (5) have been described for micro-CT data. Their dependency on specific CT scanner and/or acquisition setups impair generalizability, however. Furthermore, methods are usually...
The developed deep learning–based segmentation model was trained and validated on CT images from 1500 mice and then tested on an internal (n = 154) and external (n = 237) dataset.

On the internal test set, the model yielded a median Dice score coefficient (DSC) and Hausdorff distance (HD) of 0.984 (interquartile range [IQR], 0.977–0.988) and 0.47 mm (IQR, 0–0.51 mm), respectively.

On the external test dataset, the median DSC was 0.966 (IQR, 0.955–0.972), and the median HD was 0.31 mm (IQR, 0.30–0.32 mm).

Finally, the applicability of the model to segment lungs from high-resolution mouse micro-CT was investigated after transfer learning, with a median DSC of 0.905 and a median 95th percentile of HD of 0.33 mm.

In this study, we aimed to develop and test a deep learning–based method to automatically segment mouse lungs with varying levels of fibrosis from clinical CT images. We also explored the applicability of a transfer learning approach to obtain lung segmentation using the proposed network from high-spatial-resolution mouse micro-CT images.

### Materials and Methods

#### Study Design

A total of 323 CT images (group A; acquired at our institution between 2016 and 2017, with images used for training, validation, and testing) and 41 CT images (group B; acquired in 2014 with different resolution with regard to group A, with images used only for testing) were retrospectively collected from multiple mouse trials focused only on studying fibrosis and not on the development of a segmentation model (1,2).

The number of irradiated and healthy mice and the age ranges for the two groups are reported in Tables E1 and E2 (supplement). No further selection criteria were applied. These trials contained both healthy and irradiated (modalities: photon or particles) mice with different levels of fibrosis. Additionally, 16 CT images (group C) from healthy mice (age, 12 weeks) were retrospectively collected from an existing project from another institution (11). The CT images contained multiple mice per image, as explained later in this section. Information about the three groups is summarized in the Table. Animal care ethical approvals are reported in the respective original studies. An overview of the groups and methods is shown in Figure 1.

#### Image Acquisition

For groups A and B, CT images were acquired using a clinical PET/CT scanner (Biograph mCT, Siemens). The standard protocol used for the CT portion of PET/CT was as follows: 80 kV with 80 mAs, a pitch of 0.6 mm, section thickness of 0.6 mm, and acquisition time of 32 second (1).

For group C, CT images were acquired using a small animal microirradiator (micro-IR, X-RAD 225Cx; Precision X-Ray) with integrated micro-CT. A high-resolution 80-kVp, 2.5-mA imaging protocol filtered with an acquisition rate of 5 frames per second was used for all imaging (11).

#### Data Partition

A total of 291 CT images (90%) from group A were used for training and validation. These 291 images were split into five different folds using k-fold cross-validation (training and validation ratio of 80% and 20%, respectively). The remaining 32 CT images were used for evaluation (testing). Group B was used exclusively for testing.

A total of 80% of the CT images in group C (n = 13) were used to retrain the model using the transfer learning approach. The remaining three CT images were used for testing.

---

**Overview of the Main Parameters for the Three Mouse Groups**

| Variable                  | Group A (1,2) | Group B (1,2) | Group C (11) |
|---------------------------|---------------|---------------|--------------|
| No. of CT images          | 323           | 41            | 16           |
| No. of mice               | 1654          | 237           | 20           |
| Resolution (mm)           | 0.31 × 0.31   | 0.26 × 0.26   | 0.1 × 0.1    |
| Section thickness (mm)    | 0.6           | 0.6           | 0.1          |
| Matrix size               | 512 × 512     | 512 × 512     | 195 × 697    |
| Mouse strain              | C57BL/6       | C57BL/6       | C57BL/6      |
| Irradiation               | Photon/carbon | Photon        | None         |
| Scanner type              | Clinical CT   | Clinical CT   | Micro-CT     |
| Tube voltage (kVp)        | 80            | 80            | 80           |
| Tube current (mA)         | 80            | 80            | 2.5          |

Note.—Detailed explanation and ethical approvals of the groups are noted in the original studies in which CT was performed: Zhou et al (1), Zhou et al (2), and Granton et al (11).
Data Preparation
All the CT images were converted from Digital Imaging and Communications in Medicine format to NRRD format using the Medical Imaging Interaction Toolkit (12).

The CT images contained four to six mice each; thus, the first step was to crop each CT scan to obtain CT image data from a single mouse. Coordinates of the lung edges were identified from the reference mask and were used to find the x, y, and z coordinates to crop the corresponding CT image. This cropping resulted in a total of 1654 mice in group A (1500 mice for training and validation and 154 for testing) and 237 mice in group B. Cropped images were resized to achieve an isotropic resolution of 0.35 mm. The reference lung masks (ground truth) were segmented semiautomatically, as described by Zhou and colleagues (1,2), for both fibrotic and normal lungs. Briefly, each CT was thresholded between −900 HU and −100 HU to obtain an approximate lung segmentation. Subsequently, incorrectly assigned voxels, mostly in the trachea and primary bronchi areas, were manually removed.

The CT images from group C contained one or two mice each. Images with two mice were cropped as described previously. The total number of mice in group C was 20 (16 used to retrain the model, and four used for testing). Each cropped CT image was then resampled to an isotropic resolution of 0.2 mm. Ground truth lung masks were again segmented semiautomatically.

Finally, given the two-dimensional architecture of the proposed method, each axial section in the three-dimensional images from all groups was extracted, and the intensity was normalized between zero and 1.

Convolutional Neural Network Architecture
The convolutional neural network (CNN) was implemented with a two-dimensional U-Net architecture consisting of one encoder, which allows capturing of the global structure of the data, and one decoder, which allows a fine-grained localization. The details of the network structure are reported in Figure E1 (supplement).

The network was trained five times, each time on a different fold, using the Adam optimization algorithm and a learning rate of 0.0001. Binary cross-entropy was used as the loss function and Dice score coefficient (DSC) as the metric. The batch size was fixed at 50 for training and validation. The number of epochs was set to 30. The total model training time was approximately 15 hours on a workstation equipped with Ubuntu (version 18.04; https://releases.ubuntu.com/18.04/), as well as an Intel Xeon Processor R with eight cores and 16 GB of RAM and a GeForce GTX 1060 graphics card (NVIDIA; 6 GB).

The inference was run five times, using one different set of weights each time. The final probability lung mask was calculated as the median. Both hard thresholding (0.5) and Otsu thresholding were tested to obtain the final binary lung mask. The average inference time for one mouse was 0.31 second.

To apply this network to segment lungs from mouse CT images in group C, a transfer learning approach was implemented. In particular, the encoder was frozen during the network retraining and only the decoder was further trained; this was done to enable accurate lung segmentation.

Statistical Analysis
All the analyses described below were performed using NumPy (version 1.17.0) and MedPy (version 0.4.0), in Python (version 3.7; Python Software Foundation).

Network performance.— The DSC and the 95th percentile of the Hausdorff distance (HD) were used as metrics to quantify the segmentation accuracy, in terms of spatial overlap and...
shape mismatch (see Appendix E1 [supplement] for the definitions of DSC and HD).

**Spatial mismatch quantification.**—To find areas in the lung more sensitive to different classification, all 154 mice in the main test set were registered to a common space and then tested for deviation.

For each voxel in the difference image between the reference lung mask and the CNN segmented mask, a z-score was calculated across mice to find where the CNN segmentations more consistently deviated from the manual segmentations. See Figure E2 and Appendix E2 (supplement) for details.

**Similarity of approaches.**—To justify the use of the lung masks segmented using the proposed approach instead of those semi-automatically contoured, we compared the fibrosis indices (1,2) and the histograms of the Hounsfield units within the lungs calculated using both masks.

**Model Availability**

This pipeline is available at GitHub ([https://github.com/TransRadOnc-HIT/lung_segmentation.git](https://github.com/TransRadOnc-HIT/lung_segmentation.git)).

**Results**

**Network Performance**

The workflow described in the present analyses, with the corresponding data, is outlined in Figure 1.

A first evaluation of the proposed CNN was performed on the test dataset from group A ($n = 154$ mice). On this test set, our algorithm yielded a median DSC of 0.984, with an interquartile range (IQR) between 0.977 and 0.988 (Fig 2). The corresponding median HD (95th percentile) was 0.47 mm (IQR, 0–0.51 mm), as shown in Figure 2. The DSC and HD were also calculated using CNN lung masks binarized with a hard probability threshold of 0.5. In this case, the median DSC was 0.982 (IQR, 0.976–0.988) and the median HD was 0.47 mm (IQR, 0–0.51 mm). Given the slightly higher performance of the Otsu thresholding method, it was chosen as the binarization method.

For additional testing, the model was evaluated on an external test dataset (group B; $n = 237$ mice). Images from these mice were acquired with different resolutions as compared with those used to train the network (see Table). Observed median DSC and HD in this validation dataset were 0.966 (IQR, 0.955–0.972) and 0.31 mm (IQR, 0.30–0.32 mm), respectively (see Fig E3 [supplement]). A representative lung segmentation from the test set is shown in Figure 3; the DSC for this image was 0.981, indicating a high overlap between the reference and the CNN-segmented masks. Single discordant voxels were observed on the lung margins.

**High-Resolution Mouse Micro-CT**

Transfer learning findings are reported in Figure 4. The complete workflow is depicted in Figure 1 (right). Results for the high-resolution mice are shown in Figure 4, which shows a high overlap between the CNN and the semi-automatically segmented lung mask. For the four tested mice, the median DSC was 0.905 (IQR, 0.902–0.929) and the median 95th percentile of the HD was 0.33 mm (IQR, 2.61–2.78 mm).

**Spatial Mismatch Quantification**

To further evaluate the performance of the proposed approach, we aimed to highlight areas in the lungs that were more prone
to diverging segmentations between CNNs and reference lung masks using data from the test set of group A (n = 154). Figure 5 shows three representative views of the mouse template overlays with z-scores. As can be seen, most of the voxels classified divergently were restricted to two anatomic regions (red and blue hot spots in Fig 5).

Similarity of Approaches
Finally, data to rate similarity of semiautomatic versus CNN-based segmentation results are reported in Figure 6. Figure 6A shows the results of the gray-level histogram comparison. The two histograms showed high overlap, indicating interchangeability of the two sets of masks. Fibrosis index distributions confirmed those findings (Fig 6B).

Discussion
In this study, we present a new method for rapid and automated lung segmentation from mouse CT scans, based on a U-Net CNN. On two test datasets, the model achieved high precision, indicated by DSCs above 0.96 and 95th percentile of the HDs below 0.5 mm. These results indicate that the CNN segmentations and reference lung masks were almost perfectly overlapping, and that there were no substantial shape mismatches.

Here, the 95th percentile of the HD was used as an evaluation metric because it is widely used in other areas of research, such as in the evaluation of brain tumor segmentation (13) and brain extraction (14). The use of this metric allowed the authors to overcome the high sensitivity of the maximum HD to outliers.

In contrast to the method developed in the study by Schoppe et al (15), in which the authors presented a deep learning–based method to segment multiple organs in healthy mice, our method can be successfully applied to segment lungs from mice with different levels of fibrosis. This method can be extremely useful in translational research, such as in the analysis of radiation-induced side effects after thoracic radiation therapy. Furthermore, it can be applied to CT images with different resolutions, and not only to high-resolution micro-CT scans. A comprehensive comparison between the proposed method and other existing approaches is presented in Table E3 (supplement).
Automatic Lung Segmentation on Multiresolution CT Scans in Mice

More detailed analyses were performed to identify areas in which the CNN might yield less reliable results. Our analyses revealed that mostly voxels close to the borders of the lungs were more consistently misclassified. Of note, two hot spots of discordant segmentations were observed. In particular, the area often oversegmented by the network is close to the air-tissue interface where there is no clear separation between contiguous voxels (for example, owing to partial volume effects).

To minimize any possible registration error, a population-representing mouse template was built. This template provides a meaningful reference for the registration of mice with both fibrotic and nonfibrotic lungs. The resulting nonlinear registrations were then carefully checked, and no major failure was noticed. In contrast, selecting a random mouse from the test set as a reference yielded poor registration results, probably owing to substantial anatomic differences.

The ability of this network to segment lungs from high-resolution mouse micro-CT images was also investigated. The network was trained using a transfer learning approach on only 16 new mice. A median DSC of 0.905 was observed, which is similar to the DSC of 0.90 reported by Yan et al (5) using dynamic contrast-enhanced micro-CT images, with image resolution of 0.15 mm (isotropic), which is comparable to that of the CT images in group C (0.1 mm). The median 95th percentile HD was 0.33 mm, which is slightly higher than that for the mouse acquired with the clinical CT.

Fibrosis index and the gray-level histogram values within the lungs are often reported in studies of fibrosis development (1,2). For this reason, to justify the use of our method over semiautomatic segmentation, we extracted these two features using both masks and compared the results. Both comparisons showed high similarity, corroborating the use of this method.

Figure 4: Segmentation results for a representative high-resolution mouse. The reference lung mask (red) is overlapped with the convolutional neural network (CNN)–segmented mask (blue). The intersection between the reference lung mask and segmented masks are depicted in purple.
to segment the lungs and to use the extracted mask for any further analysis.

One limitation of this study was the two-dimensional network structure used in the current approach. This architecture inherently fails to capture global context from adjacent sections, which might explain the oversegmented spots in the mismatch analysis. This issue might be at least partially solved using a three-dimensional network architecture.

Future studies will evaluate high-resolution mouse CT data, and possibly also micro-CT, in more detail, both by including additional data in the training steps as well as by evaluating lung segmentation performance in lungs with different grades of fibrosis.

In conclusion, we developed a new method for mouse lung segmentation using a CNN. This algorithm has been tested for mice with both healthy and fibrotic lungs and yielded high DSC and low HD values. The model was successfully applied to high-resolution mouse micro-CT images.

**Author contributions:** Guarantors of integrity of entire study, P.S., M.M., C.Z., A.A., M.K.; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; approval of final version of submitted manuscript, all authors; agrees to ensure any questions related to the work are appropriately resolved, all authors; literature research, F.S., M.M., P.L., J.D., M.K.; experimental studies, M.M., C.Z., H.W., J.D., A.A.; statistical analysis, F.S., P.S., M.M., C.S., M.K.; and manuscript editing, F.S., P.S., M.M., C.S., K.R., N.B., A.K., H.W., L.D., J.D., A.A., M.K.

**Disclosures of conflicts of interest:** F.S. No relevant relationships. P.S. No relevant relationships. M.M. No relevant relationships. C.Z. No relevant relationships. C.S. No relevant relationships. K.R. No relevant relationships. N.B. No relevant relationships. A.K. No relevant relationships. H.W. Grant from Dutch Cancer Society (KWF Kankerbestrijding) to author’s institution; minority shares in Radiomics (Lieve, Belgium). L.D. No relevant relationships. P.L. Member of the external Scientific Advisory Board at CRUK Lung Cancer Center of Excellence; member of the Scientific Advisory Board (CSA) at Convert Pharmaceuticals; Chief Scientific Advisor at Oncoradiomics; member of the Advisory Board at Benelux Health Ventures SPRIL; Chief Scientific Advisor (CSA), member of the SAB at The Medical Cloud Company (Med2C); co-founder and shareholder (minority) at Oncoradiomics; co-founder and shareholder (minority) at Convert Pharmaceuticals; shareholder (minority) at The Medical Cloud Company (Med2C); shareholder (minority) at LivingMed Biotech; member of the Scientific Advisory Board at The Institute of Cancer Research; Scientific Advisor at Institut Jules Bordet, ULB; Scientific Advisor at Medeon; shareholder (minority) at Exomnis Biotech; scientific review of grants at Cancer Research UK Review Panel; consultant at Bristol Myers Squibb; Chief Scientific Advisor at Exomnis Biotech, J.D. Grants/contracts from/with Viewray, CRI – The Clinical Research Institute, Accuracy International Sarl, RaySearch Laboratories, Vision VT, Merck Serono, Astellas Pharma, AstraZeneca, Siemens Healthcare, Solution Akademie, Ergomed PLC, Surrey Research Park, Quintiles, Pharmaceutical Research Associates, Boehringer Ingelheim Pharma & CoKG, PTW-Freiburg Dr. Pychlau, Nanobiotix, Accuray, Varian; participation on a data safety monitoring board or advisory board for Merck Serono. A.A. Predict MarieCurie innovative training network (ITN), in frame of Horizon 2020 from the European Union, Marie Skłodowska-Curie grant agreement No 766276. M.K. No relevant relationships.

**Figure 5:** Spatial analysis of mismatches between the reference and the convolutional neural network (CNN)–segmented lung masks across the mice in the test set (group A). The voxels oversegmented by the CNN are colored blue, and the voxels undersegmented by the CNN are shown in red. Arrows = heart/lung interface, where the highest mismatch was observed.
Automatic Lung Segmentation on Multiresolution CT Scans in Mice

**References**

1. Zhou C, Jones B, Moustafa M, et al. Quantitative assessment of radiation dose and fractionation effects on normal tissue by utilizing a novel lung fibrosis index model. Radiat Oncol 2017;12(1):172.
2. Zhou C, Jones B, Moustafa M, et al. Determining RBE for development of lung fibrosis induced by fractionated irradiation with carbon ions utilizing fibrosis index and high-LET BED model. Clin Transl Radiat Oncol 2018;14:25–32.
3. Baiker M, Milles J, Dijkstra J, et al. Atlas-based whole-body segmentation of mice from low-contrast Micro-CT data. Med Image Anal 2010;14(6):723–737.
4. Xiao D, Zahra D, Bourgeat P, et al. Mouse whole-body organ mapping by non-rigid registration approach. In: Weaver JB, Molthen RC, eds. Proceedings of SPIE: medical imaging 2011—biomedical applications in molecular, structural, and functional imaging. Vol 7965. Bellingham, Wash: International Society for Optics and Photonics, 2011; 79650E.
5. Yan D, Zhang Z, Luo Q, Yang X. A novel mouse segmentation method based on dynamic contrast enhanced micro-CT images. PLoS One 2017;12(1):e0169424.
6. Pesapane F, Codari M, Sardanelli F. Artificial intelligence in medical imaging: threat or opportunity? Radiologists again at the forefront of innovation in medicine. Eur Radiol Exp 2018;2(1):35.
7. Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. Med Image Anal 2017;42:60–88.
8. Weng W, Zhu X. INet: Convolutional Networks for Biomedical Image Segmentation. IEEE Access 2021;9:16591–16603.
9. Milletari F, Navab N, Ahmadi SA. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. arXiv 1606.04797 [preprint] http://arxiv.org/abs/1606.04797. Posted June 15, 2016. Accessed April 2, 2019.
10. Drozdzal M, Vorontsov E, Chartrand G, Kadoury S, Pal C. The Importance of Skip Connections in Biomedical Image Segmentation. In: Carneiro G, Mateus D, Peter L, et al, eds. Deep Learning and Data Labeling for Medical Applications. DLMIA 2016, LABELS 2016. Lecture Notes in Computer Science, vol 10008. Cham, Switzerland: Springer, 2016; 179–187.
11. Granton PV, Dubois L, van Elmpt W, et al. A longitudinal evaluation of partial lung irradiation in mice by using a dedicated image-guided small animal irradiator. Int J Radiat Oncol Biol Phys 2014;90(3):696–704.
12. Wolf I, Vetter M, Wegner I, et al. The medical imaging interaction toolkit (MITK): a toolkit facilitating the creation of interactive software by extending VTK and ITK. In: Galloway RL Jr, ed. Proceedings of SPIE: medical imaging 2004—visualization, image-guided procedures, and display. Vol 5367. Bellingham, Wash: International Society for Optics and Photonics, 2004. https://doi.org/10.1117/12.535112.
13. Menze BH, Jakab A, Bauer S, et al. The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). IEEE Trans Med Imaging 2015;34(10):1993–2024.

14. Isensee F, Schell M, Pflueger I, et al. Automated brain extraction of multisequence MRI using artificial neural networks. Hum Brain Mapp 2019;40(17):4952–4964.

15. Schoppe O, Pan C, Coronel J, et al. Deep learning-enabled multi-organ segmentation in whole-body mouse scans. Nat Commun 2020;11(1):5626.