

Journal of Medical Imaging

MedicalImaging.SPIEDigitalLibrary.org

Deep learning-based segmentation of malignant pleural mesothelioma tumor on computed tomography scans: application to scans demonstrating pleural effusion

Eyjolfur Gudmundsson
Christopher M. Straus
Feng Li
Samuel G. Armato, III

SPIE.

Eyjolfur Gudmundsson, Christopher M. Straus, Feng Li, Samuel G. Armato III, "Deep learning-based segmentation of malignant pleural mesothelioma tumor on computed tomography scans: application to scans demonstrating pleural effusion," *J. Med. Imag.* 7(1), 012705 (2020), doi: 10.1117/1.JMI.7.1.012705

Deep learning-based segmentation of malignant pleural mesothelioma tumor on computed tomography scans: application to scans demonstrating pleural effusion

Eyjolfur Gudmundsson, Christopher M. Straus, Feng Li, and Samuel G. Armato III*

The University of Chicago, Department of Radiology, Chicago, Illinois, United States

Abstract. Tumor volume is a topic of interest for the prognostic assessment, treatment response evaluation, and staging of malignant pleural mesothelioma. Many mesothelioma patients present with, or develop, pleural fluid, which may complicate the segmentation of this disease. Deep convolutional neural networks (CNNs) of the two-dimensional U-Net architecture were trained for segmentation of tumor in the left and right hemithoraces, with the networks initialized through layers pretrained on ImageNet. Networks were trained on a dataset of 5230 axial sections from 154 CT scans of 126 mesothelioma patients. A test set of 94 CT sections from 34 patients, who all presented with both tumor and pleural effusion, in addition to a more general test set of 130 CT sections from 43 patients, were used to evaluate segmentation performance of the deep CNNs. The Dice similarity coefficient (DSC), average Hausdorff distance, and bias in predicted tumor area were calculated through comparisons with radiologist-provided tumor segmentations on the test sets. The present method achieved a median DSC of 0.690 on the tumor and effusion test set and achieved significantly higher performance on both test sets when compared with a previous deep learning-based segmentation method for mesothelioma. © 2020 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: [10.1117/1.JMI.7.1.012705](https://doi.org/10.1117/1.JMI.7.1.012705)]

Keywords: tumor segmentation; deep learning; computed tomography; malignant pleural mesothelioma; convolutional neural networks.

Paper 19144R received Jun. 23, 2019; accepted for publication Dec. 24, 2019; published online Jan. 29, 2020.

1 Introduction

Malignant pleural mesothelioma is a cancer of the pleura, the tissue that forms the inner lining of the thoracic cavity and the outer lining of the lungs. Mesothelioma is an aggressive disease with a median survival time of 13–16 months for patients on standard chemotherapeutic treatment; a patient history of asbestos exposure is evident in ~80% of cases.^{1,2}

Mesothelioma commonly grows as rind-like pleural thickening and exhibits an irregular, non-spherical morphology.^{3,4} A majority of mesothelioma patients present with, or develop, pleural effusion.^{4,5} These aspects of the disease, combined with the often large anatomical extent of the tumor and low contrast between tumor and adjacent soft tissues, complicate the image-based evaluation of tumor bulk for treatment response assessment in mesothelioma patients. Tumor bulk is clinically assessed during treatment through summed linear thickness measurements made according to the modified Response Evaluation Criteria in Solid Tumors guidelines.^{6,7} Considerable observer variability in mesothelioma tumor thickness measurements and the irregular morphology and growth patterns of this disease have called into question how accurately such measurements capture the extent of disease at a given treatment time point.^{8,9} Volumetric segmentation of tumor may provide a more comprehensive assessment of tumor bulk; however, the time-consuming nature of manual outlining of the tumor has prevented the use of volumetric assessment for clinical purposes. A number of studies have investigated the correlation of image-based mesothelioma tumor volume with patient survival and tumor response to treatment and the potential role of tumor volume in the staging of this disease.^{10–14}

*Address all correspondence to Samuel G. Armato III, E-mail: s-armato@uchicago.edu

Deep convolutional neural networks (CNNs) are machine learning-based classifiers that have shown promise for the classification and segmentation of lesions and other anatomy in medical images.^{15,16} Deep CNNs consist of layers of convolutional filters that, through an optimization process, learn to detect features in images that are associated with a given classification of the images; downsampling the input image allows the network to analyze and combine features at different scales of the image. A common CNN architecture used for the segmentation of structures on medical images is the U-Net architecture.¹⁷

Data scarcity is a common issue in the training of deep CNNs for medical imaging-related tasks. To alleviate this problem, researchers have investigated the use of convolutional layers pretrained on large natural image datasets, such as ImageNet,¹⁸ to initialize deep CNNs applied to clinical tasks through a process called “transfer learning.”¹⁹ This strategy can be used to directly extract feature descriptors of medical image regions-of-interest for further classification using a nondeep learning-based machine learning classifier (e.g., in the classification of benign and malignant lesions) or to initialize the layers of a deep CNN that is subsequently fine-tuned for a given clinical task through further training.^{20,21}

Previous work on the automated segmentation of mesothelioma tumor on CT scans includes a stepwise method that aimed to isolate pleural thickening in patients through the segmentation of the lungs and the ribcage²² and our study that used a deep CNN-based segmentation method based on the U-Net architecture for the automated segmentation of mesothelioma on CT scans (“2018 Method”).²³ In our prior deep learning-based study, the 2018 Method showed superior performance to the nondeep learning-based method on an independent test set of CT scans of mesothelioma patients with radiologist-provided tumor contours. It was found, however, that the 2018 Method showed increased disagreement with observer-provided tumor contours in the presence of pleural effusion.

Other studies that have explored the correlation of mesothelioma tumor volume with patient survival and other clinically relevant factors have used semiautomated methods to obtain image-based patient tumor volumes. These methods typically rely on user-provided tumor contours on every few axial CT sections; interpolation is subsequently used to obtain tumor segmentations on other sections.^{12,13} Such semiautomated methods are more time-efficient when compared with the fully manual approach; yet, they are still too time consuming to be adapted for day-to-day clinical purposes.

This study aimed to improve the deep learning-based segmentation of malignant pleural mesothelioma tumor on CT scans of patients who present with pleural effusion through the initialization of deep CNNs with pretrained convolutional layers, an expanded set of effusion-containing CT scans for training of the networks, and a comprehensive strategy for the evaluation of the segmentation performance of the networks during training. These modifications of the deep CNN-based method for the segmentation of mesothelioma were intended to aid in the reduction of nontumor pixels (in particular, pleural effusion) erroneously classified as tumor by the networks. The development of a robust automated segmentation method of mesothelioma tumor should focus on the exclusion of effusion from segmented tumor, and the performance of such a method should be assessed on a set of scans that exhibit both tumor and pleural fluid. Volumetric segmentations of mesothelioma tumor could provide clinicians with additional data to aid them in their patient management-related decisions; a deep learning-based automated segmentation method for mesothelioma could eventually lead to an efficient way to estimate tumor bulk for the clinical evaluation of treatment response of this aggressive disease and allow for efficient data collection toward advanced pixel-based analyses of tumor in research studies.

2 Materials and Methods

Mesothelioma patients commonly present with pleural effusions and atelectatic (i.e., collapsed) lung; these abnormalities have considerable overlap in Hounsfield unit (HU) values with mesothelioma.²⁴ Mesothelioma may invade the chest wall, mediastinum, and/or the abdomen in late-stage patients.²⁵ This disease rarely affects both sides of the chest; ~95% of mesothelioma patients exhibit unilateral disease.²⁶ The present study addressed the identification of mesothelioma tumor in patients who exhibited unilateral diseases that had not invaded other structures

or organs. Deep CNNs were trained separately for the segmentation of disease in the left and right hemithoraces. Results of the present method were compared with (1) a set of manual tumor outlines constructed by three thoracic radiologists on CT sections of two test sets not included in the training set and (2) the output of the 2018 Method on the same test sets.²³

2.1 Data Preprocessing

All CT scans used for training, validation, and testing underwent an in-house thoracic segmentation method developed in MATLAB (Mathworks Inc., Natick, Massachusetts) to exclude the patient table and surrounding air. All CT sections used for training, validation, and testing were converted from HU values to 32-bit floating-point values in the range [0,1] according to the following piecewise linear scaling:

- Pixels outside the thorax and pixels of value equal to or below -1000 HU were assigned a value of 0, and pixels of value ≥ 240 HU were assigned a value of 1.
- Pixel values on the range $(-1000 \text{ HU}, -700 \text{ HU}]$ were linearly scaled to the floating-point range $(0, 0.15]$.
- Pixel values on the range $(-700 \text{ HU}, -160 \text{ HU}]$ were linearly scaled to the floating-point range $(0.15, 0.25]$.
- Pixel values on the range $(-160 \text{ HU}, 240 \text{ HU})$ were linearly scaled to the floating-point range $(0.25, 1)$.

This rescaling of pixel values was used to increase the contrast of pixels of values in the “soft-tissue window” range of $[-160 \text{ HU}, 240 \text{ HU}]$. The values of the HU ranges and the corresponding floating-point ranges were determined prior to the study through visualization of a subset of the training set. Following the conversion to the range $[0, 1]$, the rescaled arrays were copied to three-channel arrays and appropriately normalized for use with the pretrained networks of this study.

2.2 Training Set

The training set of the present study consisted of 2663 sections from 76 scans of 61 mesothelioma patients with disease in the left hemithorax and 2567 sections from 78 scans of 65 mesothelioma patients with disease in the right hemithorax (see Table 1). Of these sections, 525 and 520 sections presented with pleural effusion in the left and right hemithorax, respectively. CT sections that did not present with tumor were excluded from the training set of the present study.

2.3 Test Sets

Two test sets of CT sections of mesothelioma patients with radiologist-provided reference tumor segmentations were used for testing the deep CNNs trained in this study. The first of these test

Table 1 Characteristics of CT scans used for training the deep CNNs.

Characteristic	Value	
	Left hemithorax	Right hemithorax
No. of patients	61	65
No. of CT scans	76	78
No. of CT sections	2663	2567
Median slice thickness (range)	3 (0.625–5) mm	3 (0.625–5) mm
Median pixel spacing (range)	0.731 (0.561–0.977) mm	0.720 (0.588–0.943) mm

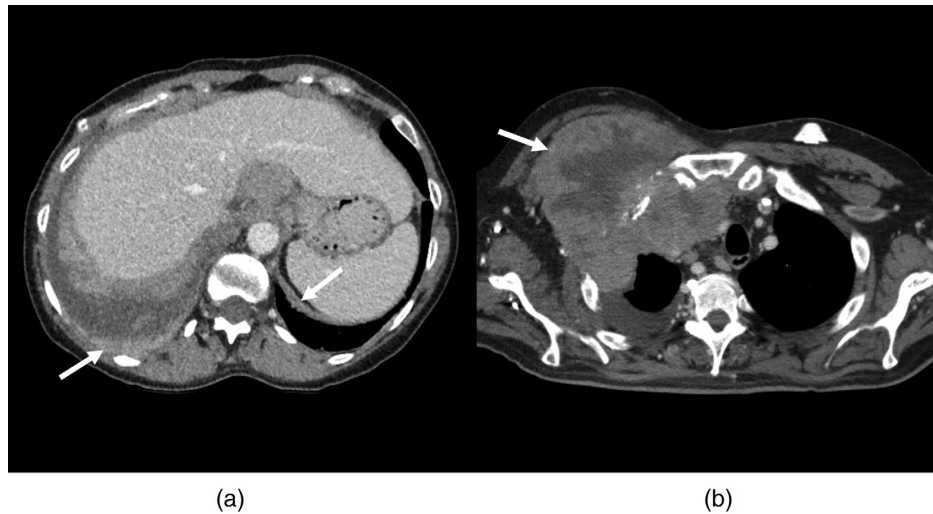


Fig. 1 Examples of CT sections excluded from the test sets of this study due to (a) bilateral tumor (tumor foci shown with white arrows) and (b) tumor chest wall invasion (chest wall invasion shown with white arrow).

sets consisted solely of sections on which both tumor and pleural effusion were present; this test set was collected specifically for this study to test the segmentation performance of the present method and the 2018 Method on a set of cases with a variety of pleural effusion presentations. The second test set consisted of sections that did not all present with both tumor and pleural effusion; this set was used in the present study to test the segmentation performance of the present method and the 2018 Method on a more general collection of CT scans of mesothelioma patients.

The test set specifically created for the present study (“Tumor and Effusion Test Set”) consisted of 94 axial CT sections (that all exhibited both tumor and pleural effusion) randomly selected from 46 CT scans of 34 patients not included in the training or validation sets of the present study. Reference tumor segmentations on this test set were constructed by a radiologist experienced in the measurement of mesothelioma (F.L.). Sections exhibiting bilateral disease and invasion of anatomic structures adjacent to the pleura were excluded from the test set (see Fig. 1). To reduce anatomic correlation between sections from the same scan, only a single section randomly selected from within each of the upper, mid, and lower thoracic regions of any one scan were included in the test set (for a maximum of three sections per scan); furthermore, all sections from the same scan were separated in the axial direction by at least 1 cm. Of the 94 axial sections of this test set, 40 had left-hemithorax disease and 54 had right-hemithorax disease (see Table 2).

The second test set (“Test Set 2”) of the present study consisted of 130 CT sections from 43 scans of 43 mesothelioma patients not included in the training or validation sets of the present

Table 2 Test sets used for segmentation performance assessment.

Characteristic	Value	
	Tumor and Effusion Test Set	Test Set 2
No. of sections (%)		
Left-sided disease	40 (43)	40 (31)
Right-sided disease	54 (57)	90 (69)
Median slice thickness (range)	3 (2.5–4) mm	3 (1–5) mm
Median pixel spacing (range)	0.747 (0.557–0.926) mm	0.723 (0.535–0.883) mm

study. This test set was created through the combination of the two test sets of our prior published study on the deep learning-based segmentation of mesothelioma (i.e., the “2018 Method”).²³ To simplify the analysis of the present study, a single set of reference tumor contours for Test Set 2 was constructed using only the reference contours provided by the single observer with the highest average interobserver Dice similarity coefficient (DSC) value on each of the test sets of our prior study (observers A and 4 of our prior study). Exclusion criteria for Test Set 2 included prominent calcifications, chest-wall invasion, surgical intervention, a mean DSC value across all observers ≤ 0.5 , and lack of observer agreement on the laterality of disease. Of the 130 axial sections in Test Set 2, 40 exhibited left-hemithorax disease and 90 exhibited right-hemithorax disease (see Table 2). In this test set, 39 sections (30%) exhibited pleural effusion.

2.4 Deep CNN Architecture

The Visual Geometry Group (VGG) 16 network architecture, pretrained on the ImageNet database, was used as the downsampling path of a U-Net deep CNN for the experiments of the present study.^{17,27–29} The layers of the downsampling path were initialized using weights acquired when the VGG16 was trained with scale-jittering on the ImageNet dataset of natural images (configuration D in the original paper of Simonyan et al.).^{18,27} The weights of the first two pretrained convolutional layers of the network were kept fixed during training; other pretrained layers were fine-tuned during training. The weights of the layers of the upsampling path of the networks were initialized using Glorot uniform initialization.³⁰

The network architecture is shown in Fig. 2. The network accepted as input a 512×512 image matrix and produced a tumor segmentation mask of the same size as the input. Convolutional layers in the network were followed by a rectified linear unit (ReLU) activation function, except at the last layer, for which a sigmoid activation function was used to produce values on the range (0, 1).³¹ The continuous output of the network was transformed into binary tumor segmentations using a predetermined threshold value of 0.5. A 2×2 max pooling operation with stride 2 was used to implement the downsampling of the feature matrix at each level of the downsampling path. The number of feature channels was doubled at each downsampling step, starting with 64 channels at the input level of the network. Dropout layers of probability 0.5 were used during later stages of the downsampling path to prevent overfitting.³² At each level of the upsampling path, a two-dimensional upsampling operation using nearest-neighbor interpolation was applied to the feature matrix, and the resulting feature map was concatenated with the feature map from the corresponding level of the downsampling path.

Network loss during training was calculated per image as the binary cross-entropy L , summed over all pixels, between the deep CNN-predicted segmentation and the corresponding reference tumor segmentation:

$$L(t_i, p_i) = -[t_i \log(p_i) + (1 - t_i) \log(1 - p_i)] \quad (1)$$

where t_i is an indicator variable taking the value 1 if the reference classification of pixel i is tumor and 0 otherwise, and p_i is the deep CNN-predicted probability that pixel i is tumor or background. The Adam method was used to optimize the network during training using an initial learning rate of 10^{-5} , chosen after initial investigations on a subset of the training set.³³ The deep CNN architecture was implemented with the Keras and Tensorflow deep learning frameworks.³⁴ Experiments were run using a batch size of 1 on a scientific computing cluster at The University of Chicago using Nvidia GeForce GTX Titan and Nvidia Tesla K20c Kepler-class graphics processing units (GPUs; Nvidia, Santa Clara, California).

2.5 Experiments

Deep CNNs were trained separately on sections and reference segmentations of mesothelioma patients with visible disease in the left and right hemithoraces. Validation sets were used to select the optimal deep CNN of each hemithorax for application to the test sets. For the left hemithorax, 372 CT sections from 10 scans of nine patients were excluded from the training set and used as

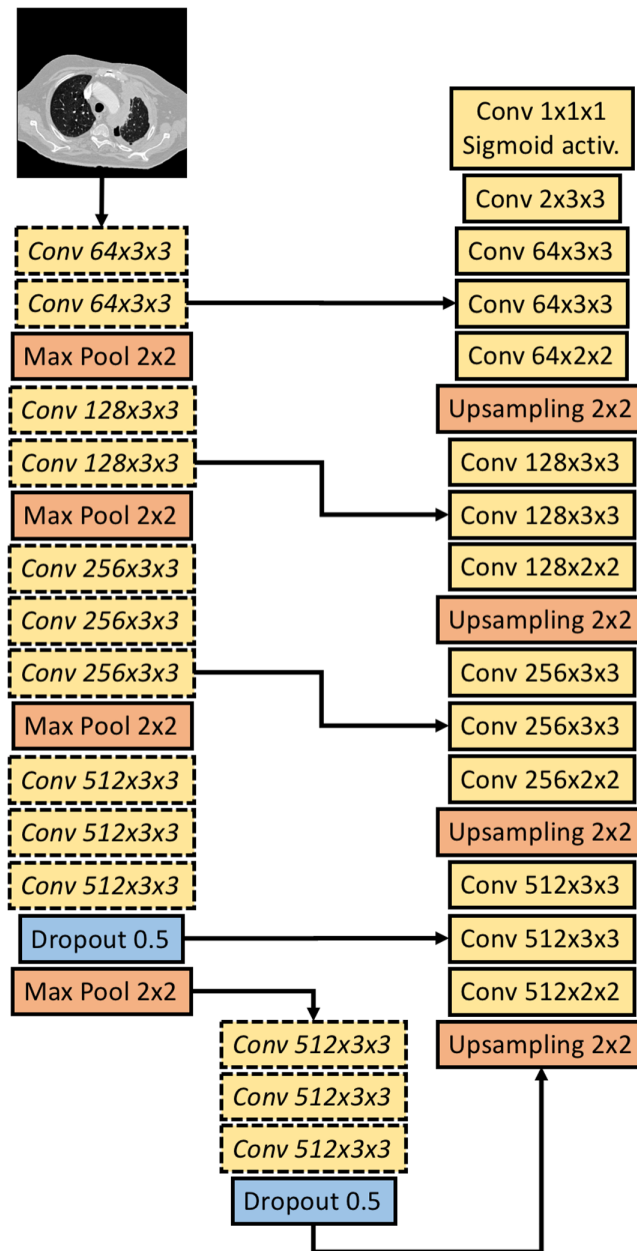


Fig. 2 The VGG16/U-Net deep CNN architecture of the present study. Dashed outlines indicate VGG16 layers pretrained on the ImageNet database. Arrows indicate the flow of the input matrix and feature matrix through the network. Convolutional layers are labelled with “Conv” and a triplet of numbers that represents the number of feature channels, height of the convolution window, and width of the convolution window, respectively. A dropout probability of 0.5 was used during training. ReLU activation functions were used following convolutional layers, except where noted. Upsampling was implemented through nearest neighbor interpolation.

a validation set during training; for the right hemithorax, 316 CT sections from 10 scans of 10 patients were excluded from the training set and used as a validation set during training.

The training and validation sets of each hemithorax were divided into subsets of (1) sections that exhibited tumor with no apparent effusion (“tumor only”) and (2) sections that exhibited both tumor and effusion (“tumor and effusion”). Table 3 presents the number of sections in each subset of the training and validation sets of each hemithorax. The training set of each side of the chest contained approximately four times as many sections that exhibited tumor without apparent effusion than sections that exhibited both tumor and effusion. To determine the optimal relative proportion of the two classes of sections (i.e., “tumor only” and “tumor and effusion”) in

Table 3 Division of training and validation sets of each hemithorax into sections that exhibit tumor without apparent effusion (“tumor only”) and sections that exhibit tumor with apparent effusion (“tumor and effusion”). No scans of the same patient formed a part of both the training and validation sets of a given hemithorax.

Characteristic	No. of scans	Median no. of sections per scan
Left hemithorax, training set		
Tumor only	33 (2138 sections)	58 (range: 8–126)
Tumor and effusion	54 (525 sections)	7 (range: 1–68)
Left hemithorax, validation set		
Tumor only	4 (275 sections)	70 (range: 58–77)
Tumor and effusion	8 (97 sections)	9 (range: 3–33)
Right hemithorax, training set		
Tumor only	35 (2047 sections)	59 (range: 14–112)
Tumor and effusion	50 (520 sections)	6 (range: 1–48)
Right hemithorax, validation set		
Tumor only	4 (215 sections)	56 (range: 16–88)
Tumor and effusion	8 (101 sections)	6.5 (range: 1–38)

the training set of each hemithorax, the values of 1:1, 2:1, and 4:1 were explored for the relative proportion of the two classes during training; the validation set of each hemithorax was used to determine the optimal relative frequency of the two classes.

Three metrics of segmentation performance were calculated on the validation set of each hemithorax during training: the median DSC, the median average Hausdorff distance (AHD; also known as the modified Hausdorff distance), and the ratio P of the total number of predicted tumor pixels to the total number of reference tumor pixels. The DSC is defined as follows:

$$\text{DSC}(S_1, S_2) = \frac{2|S_1 \cap S_2|}{|S_1| + |S_2|}, \quad (2)$$

where $|S_1|$ and $|S_2|$ represent the respective area of each segmentation and $|S_1 \cap S_2|$ represents the area of the intersection of the two segmentations.^{35,36} AHD is an evaluation metric for a pair of segmentations that takes into account the spatial location of the segmentation boundaries:^{36,37}

$$\text{AHD}(A, B) = \max [d(A, B), d(B, A)] \quad (3)$$

where A and B represent the two tumor contours (i.e., tumor segmentation outlines) under comparison and $d(A, B)$ is defined as the Euclidean distance to the nearest point on contour B from a point on contour A averaged across all points on contour A . The AHD metric has been found to be more robust in the presence of outliers relative to the original Hausdorff distance.³⁷ These three metrics were calculated on the validation set approximately every 1000 network updates during training. Deep CNNs were trained for 3×10^5 updates or until performance metric values on the validation set indicated that the deep CNN had started to overfit the training set.

Minimal data augmentation was applied to the training set due to the inherent asymmetry of chest anatomy. For the present study, a rotation of either -10 deg or $+10$ deg and scaling of either 0.9 or 1.1 were selected for each CT scan of the training set. The values of rotation and scaling were determined by visualizing different rotation angles and scaling values on example CT sections from the training set.

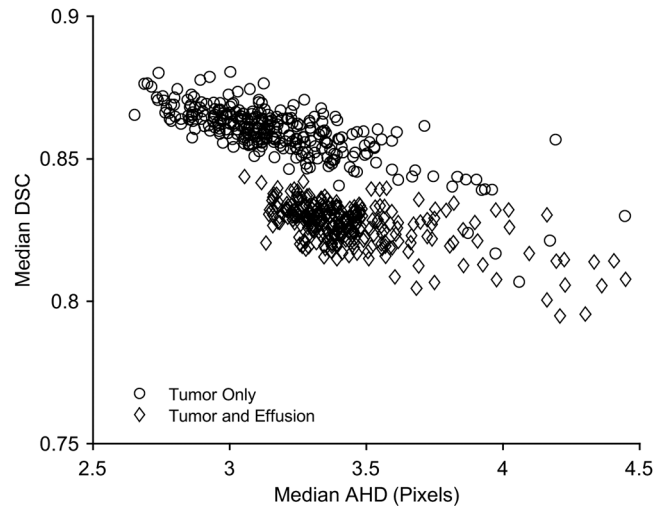


Fig. 3 Median DSC as a function of the median AHD obtained on the “tumor only” and “tumor and effusion” subsets of the validation set during training, shown for the left hemithorax and a 4:1 relative frequency of tumor only and tumor and effusion sections in the training set. The same general trend was observed for both sides of the chest across all relative frequencies of tumor only and tumor and effusion sections in the training sets.

Median AHD was found to be a more representative metric of overall segmentation performance during training than the median DSC. Figure 3 shows a scatter plot of the median DSC and median AHD values obtained for the “tumor only” and “tumor and effusion” subsets of the validation set of the left hemithorax. The general trend shown in Fig. 3, for which lower median AHD values were associated with relatively high median DSC values and higher median DSC values were not necessarily associated with low median AHD values, was evident on plots of median DSC as a function of median AHD for the validation sets of both hemithoraces. The minimum median AHD on the validation set of each hemithorax was therefore selected as the segmentation performance metric for the selection of CNNs for application to the test sets. Only a single optimal deep CNN was chosen for application to the test sets for each hemithorax.

To evaluate interobserver agreement in the manual segmentation of mesothelioma tumor on CT sections that exhibit pleural effusion, the mean interobserver DSC value among five radiologists (i.e., the mean of the ten interobserver comparisons made for each section) on a set of 69 axial CT sections from 27 scans of 27 patients was compared between (1) sections for which at least one observer excluded an area of pleural effusion from the tumor contours and (2) sections for which none of the five observers excluded an area of pleural effusion from tumor contours. These images formed a part of Test Set 2 of the present study and were used in a previously published study on observer variability in mesothelioma tumor measurements.³⁸

2.6 Statistical Analysis

The two-sided Wilcoxon signed-rank test was used to test the null hypothesis that the distributions of DSC values and AHD values were identical for the present method and the 2018 Method when compared with reference tumor segmentations of the two test sets of this study. The two-sided Wilcoxon rank-sum test was used to test the null hypothesis that the distributions of average interobserver DSC values were identical for (1) sections on which at least one out of five radiologists excluded an area of effusion from mesothelioma tumor contours and (2) sections on which none of the five radiologists excluded an area of effusion from tumor contours on a set of 69 CT sections. The Bonferroni–Holm correction was used to account for the five statistical comparisons made in this study.³⁹ Statistical tests were made using MATLAB.

Bland–Altman plots were used (1) to evaluate the agreement between tumor area segmented by the present method and observer-segmented tumor area on each test set of this study and (2) to evaluate the agreement between tumor area segmented by the 2018 Method and observer-segmented tumor area on each test set of this study. Absolute differences in the segmented area of

the computerized methods and observer-segmented area were found to correlate with the average segmented tumor area of the segmentation approaches being compared, violating the normality assumption for calculation of 95% limits of agreement according to the Bland–Altman method. Therefore, the 95% limits of agreement were estimated using relative differences in segmented area as $d \pm 1.96s$, where d is the mean and s is the standard deviation of the relative differences between the two segmentation approaches being compared.⁴⁰ The standard error (SE) of d was estimated as $\sqrt{s^2/n}$ and the SE of the 95% limits of agreement was estimated as $\sqrt{3s^2/n}$, where n is the number of samples. 95% confidence intervals (CIs) for d and the 95% limits of agreement were found by adding and subtracting $1.96 \times \text{SE}$ from each value in question.⁴¹

3 Results

For the left hemithorax, a 4:1 relative frequency of tumor only and tumor and effusion subsets of the training set was found to achieve the lowest median AHD on the validation set. For the right hemithorax, a 1:1 relative frequency of tumor only and tumor and effusion subsets of the training set was found to achieve the lowest median AHD on the validation set. All results presented in this section refer to these cases for the respective side of the chest.

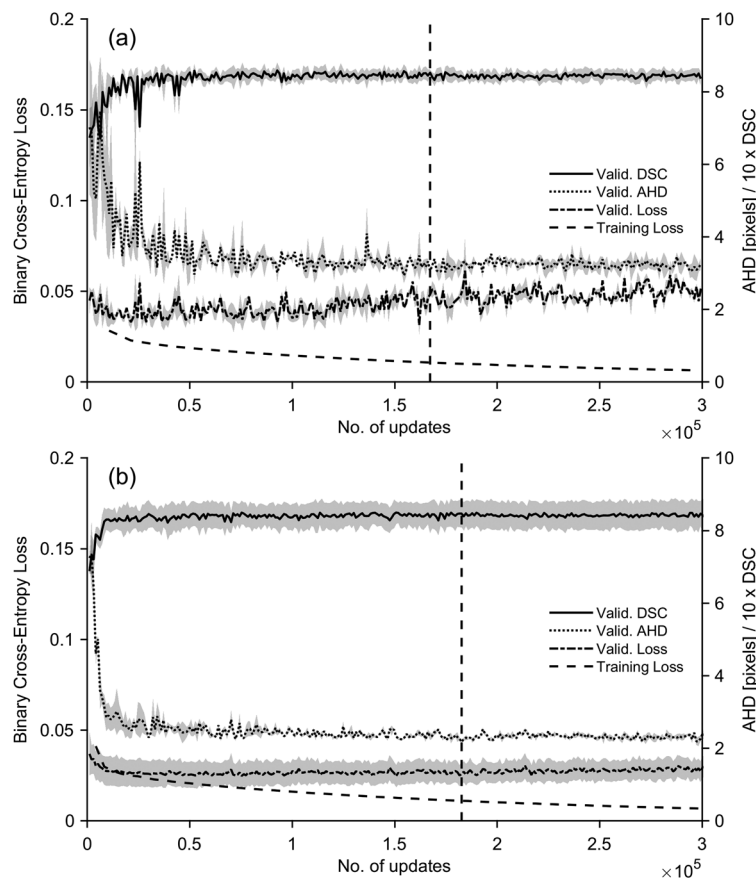


Fig. 4 Average loss L on the training set, and average loss L , median DSC and median AHD on the validation sets during training (a) of the left-hemithorax deep CNN and (b) of the right-hemithorax deep CNN. Solid lines indicate average values on the training set across the tumor only and tumor and effusion validation sets of each hemithorax. Shaded areas indicate the range of the average loss, median DSC, and median AHD across the tumor only and tumor and effusion validation sets of each hemithorax. Median DSC values are shown with a scaling factor of 10 for visual clarity. Validation set performance was assessed approximately every 1000 updates. The vertical dashed lines indicate the training updates after which the deep CNNs were applied to the test sets.

Table 4 Minimum median AHD value achieved on the tumor and effusion validation set of the left hemithorax and the minimum median AHD value achieved on the tumor only validation set of the right hemithorax and the corresponding average loss L , median DSC value, median AHD value, and ratio of the number of predicted tumor pixels to reference tumor pixels (pixel ratio) for both validation sets of each hemithorax during training.

Hemithorax	Validation set	Metric	Training update	Value
Left	Tumor and effusion	Minimum median AHD	167,236	3.05 pixels
		Average L	—	0.048
		Median DSC	—	0.844
		Pixel ratio	—	0.95
	Tumor Only	Median AHD	—	3.07 pixels
		Average L	—	0.048
		Median DSC	—	0.844
		Pixel ratio	—	0.82
Right	Tumor and effusion	Median AHD	182,620	2.22 pixels
		Average L	—	0.032
		Median DSC	—	0.885
		Pixel ratio	—	1.08
	Tumor only	Minimum median AHD	—	2.12 pixels
		Average L	—	0.020
		Median DSC	—	0.809
		Pixel Ratio	—	1.00

3.1 Training

The average binary cross-entropy loss L on the training and validation set of each hemithorax and the median DSC and median AHD across the validation set of each hemithorax during training are shown in Fig. 4 for each hemithorax. The lines in Fig. 4 indicate the average loss on the training set, and the average loss, average median DSC, and average median AHD across the tumor only and tumor and effusion validation sets for each hemithorax. The shaded areas in Fig. 4 indicate the range of the loss, median DSC, and median AHD between the tumor only and tumor and effusion validation sets of each hemithorax. An overall better segmentation performance across both subsets of the validation sets of each hemithorax was found when choosing the optimal network based on the minimum median AHD achieved on the tumor and effusion subset of the left hemithorax and based on the minimum median AHD achieved on the tumor only subset of the right hemithorax. Table 4 lists the minimum median AHD achieved on the tumor and effusion and the tumor only validation set of the left and right hemithorax, respectively, and the corresponding median DSC, loss, and ratio of the number of deep CNN-predicted tumor pixels to reference tumor pixels for both validation sets at the corresponding training update.

3.2 Tumor and Effusion Test Set

Figure 5 shows boxplots of DSC values and AHD values obtained when comparing the predicted tumor segmentations of the present deep CNN method and the 2018 Method with the reference tumor contours on the Tumor and Effusion Test Set. The median DSC and median AHD for the

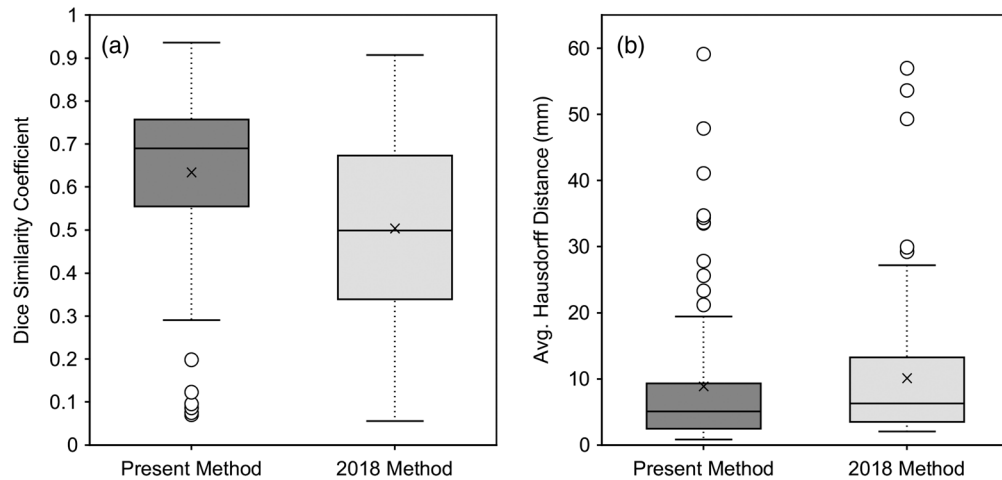


Fig. 5 Boxplots showing segmentation performance of the present method and the 2018 Method when comparing predicted tumor segmentations with radiologist-acquired reference tumor segmentations on the Tumor and Effusion Test Set. Horizontal lines inside boxes indicate the median value of each distribution; crosses indicate the mean value of each distribution. (a) DSC values on the Tumor and Effusion Test Set. (b) AHD values on the Tumor and Effusion Test Set.

present method on this test set were 0.690 (range: 0.070–0.936) and 5.1 mm (range: 0.9–59.1 mm), respectively. The median DSC and median AHD for the 2018 Method on the same CT sections were 0.499 (range: 0.055–0.907) and 6.3 mm (range: 2.0–57.0 mm), respectively. Differences in the distributions of DSC values ($p < 0.00001$) and AHD values ($p < 0.001$) between the two methods on this test set were found to be statistically significant using the two-sided Wilcoxon signed-rank test.

Figure 6(a) shows a Bland–Altman plot of the relative differences in the segmented tumor area by the present deep CNN method and the observer-segmented tumor area on the Tumor and Effusion Test Set. The mean relative difference in the segmented tumor area between the present method and the observer-segmented area was -8.2% (95% CI: -17.5% to 1.1%) with 95% limits of agreement $[-96.3\%, 79.9\%]$ (95% CIs: -112.1% to -80.6% and 64.2% to 95.6% for the lower and upper limits, respectively). Figure 6(b) shows a Bland–Altman plot of the relative differences in the segmented tumor area by the 2018 Method and the tumor area segmented by the observer on the Tumor and Effusion Test Set. The mean relative difference in the

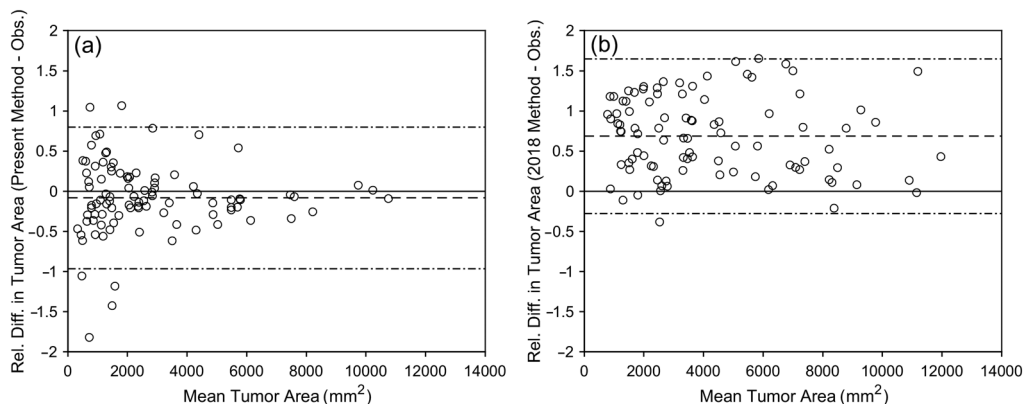


Fig. 6 Bland–Altman plots showing relative differences between the segmented tumor area of the present method and of the 2018 Method and the observer-segmented tumor area on the Tumor and Effusion Test Set. Means of relative differences and 95% limits of agreement are shown as dashed lines. (a) Relative differences in tumor area between the present method and the observer on the Tumor and Effusion Test Set. (b) Relative differences in tumor area between the 2018 Method and the observer on the Tumor and Effusion Test Set.

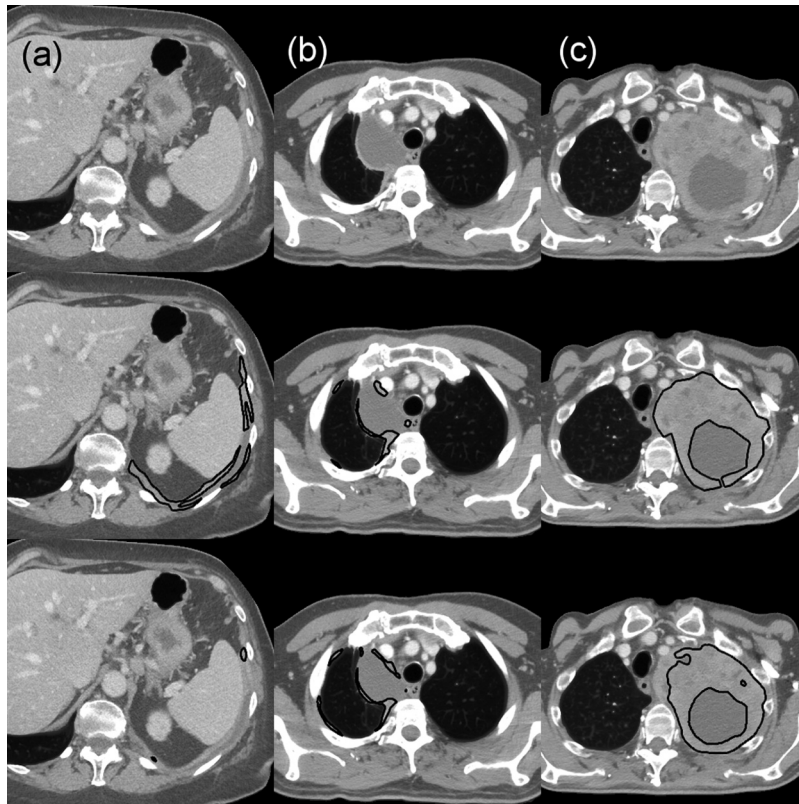


Fig. 7 Preprocessed CT sections (top), observer reference tumor segmentations (middle; black outlines), and predicted tumor segmentations by the present method (bottom; black outlines), for three sections from different CT scans of the Tumor and Effusion Test Set. Sections were selected at random from (a) the bottom 10th percentile (DSC = 0.086, AHD = 33.6 mm), (b) the interquartile range (DSC = 0.619, AHD = 4.5 mm), and (c) the top 10th percentile (DSC = 0.880, AHD = 3.0 mm) of the DSC values obtained when comparing predicted tumor segmentations of the present method and observer reference segmentations.

segmented tumor area between the 2018 Method and the observer-segmented area was 68.6% (95% CI: 58.4%–78.7%) with 95% limits of agreement [−27.7%, 164.9%] (95% CIs: −44.9% to −10.5% and 147.7% to 182.1% for the lower and upper limits, respectively).

Figure 7 shows the preprocessed CT sections, reference tumor segmentations, and predicted tumor segmentations of the present method for three example CT sections selected at random from the lowest 10th percentile, the interquartile range, and the top 10th percentile of the DSC values found when comparing predicted tumor segmentations of the present method with reference segmentations on the Tumor and Effusion Test Set of this study.

3.3 Test Set 2

Figure 8 shows boxplots of DSC values and AHD values obtained when comparing the predicted tumor segmentations of the present method and the 2018 Method with the set of reference tumor contours on Test Set 2. The median DSC and median AHD for the present method on Test Set 2 was 0.780 (range: 0.175–0.927) and 2.9 mm (range: 0.7–50.7 mm), respectively. The median DSC and median AHD for the 2018 Method on Test Set 2 was 0.764 (range: 0.108–0.938) and 3.3 mm (range: 0.8–56.9 mm), respectively. The difference in the distributions of AHD values between the two methods on this test set was found to be statistically significant ($p = 0.008$) using the two-sided Wilcoxon signed-rank test; the difference in the DSC value distributions did not reach statistical significance on this test set using the two-sided Wilcoxon signed-rank test ($p = 0.23$).

Figure 9(a) shows a Bland–Altman plot of the relative differences in the segmented tumor area by the present method and the average observer-segmented tumor area on Test Set 2.

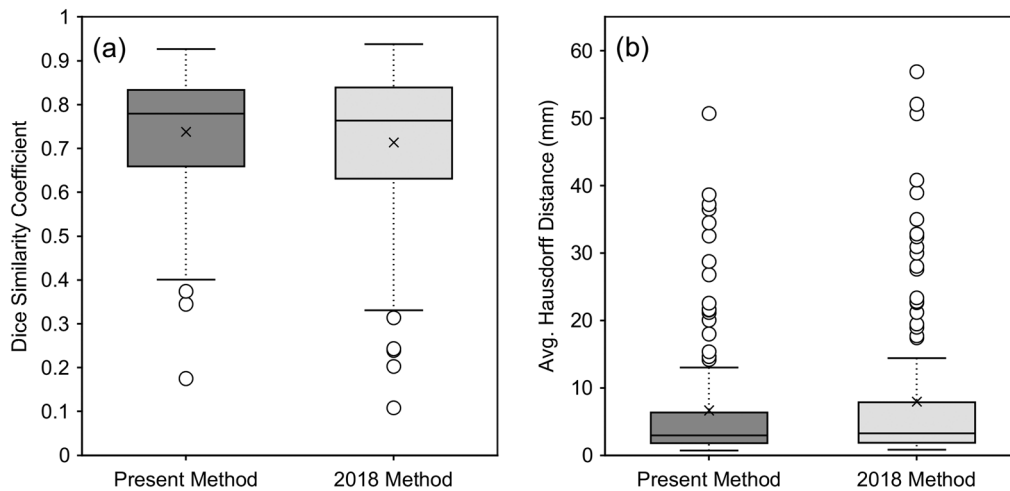


Fig. 8 Boxplots showing segmentation performance of the present method and the 2018 Method when comparing predicted tumor segmentations with radiologist-acquired reference tumor segmentations on Test Set 2. Horizontal lines inside boxes indicate the median value of each distribution; crosses indicate the mean value of each distribution. (a) DSC values on Test Set 2. (b) AHD values on Test Set 2.

The mean relative difference in the segmented tumor area by the present method and the observer-segmented area was -17.4% (95% CI: -23.1% to -11.7%) with 95% limits of agreement $[-80.8\%, 46.0\%]$ (95% CIs: -90.5% to -71.2% and 36.4% to 55.6% for the lower and upper limits, respectively). Figure 9(b) shows a Bland–Altman plot of the relative differences in the segmented tumor area by the 2018 Method and the observer-segmented tumor area on Test Set 2. The mean relative difference in the segmented tumor area between the 2018 Method and the observer-segmented area on this test set was 11.8% (95% CI: 3.3% – 20.2%) with 95% limits of agreement $[-83.1\%, 106.6\%]$ (95% CIs: -97.5% to -68.7% and 92.2% to 121.0% for the lower and upper limits, respectively).

Figure 10 shows the preprocessed CT sections, reference tumor segmentations, and predicted tumor segmentations of the present method for three example CT sections selected at random from the lowest 10th percentile, the interquartile range, and the top 10th percentile of the DSC values found when comparing predicted tumor segmentations of the present method with reference segmentations on Test Set 2 of this study.

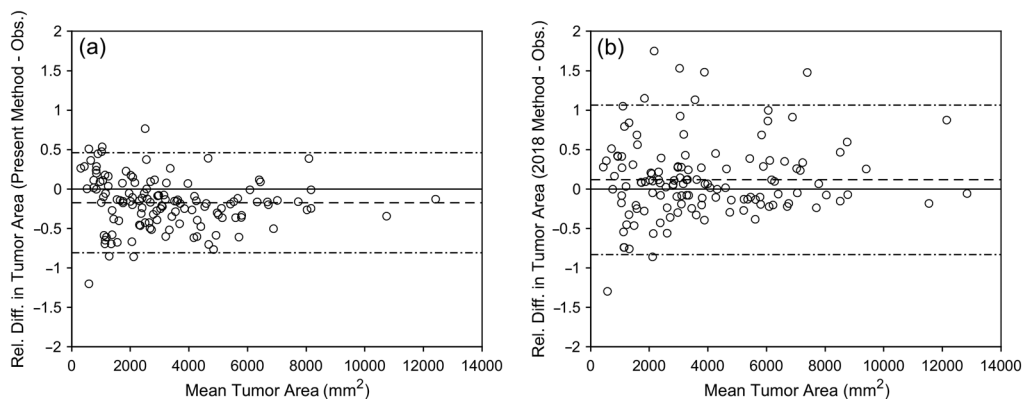


Fig. 9 Bland–Altman plots showing relative differences between the segmented tumor area of the present method and of the 2018 Method and the observer-segmented tumor area on Test Set 2. Means of relative differences and 95% limits of agreement are shown as dashed lines. (a) Relative differences in tumor area between the present method and the observer on Test Set 2. (b) Relative differences in tumor area between the 2018 Method and the observer on Test Set 2.

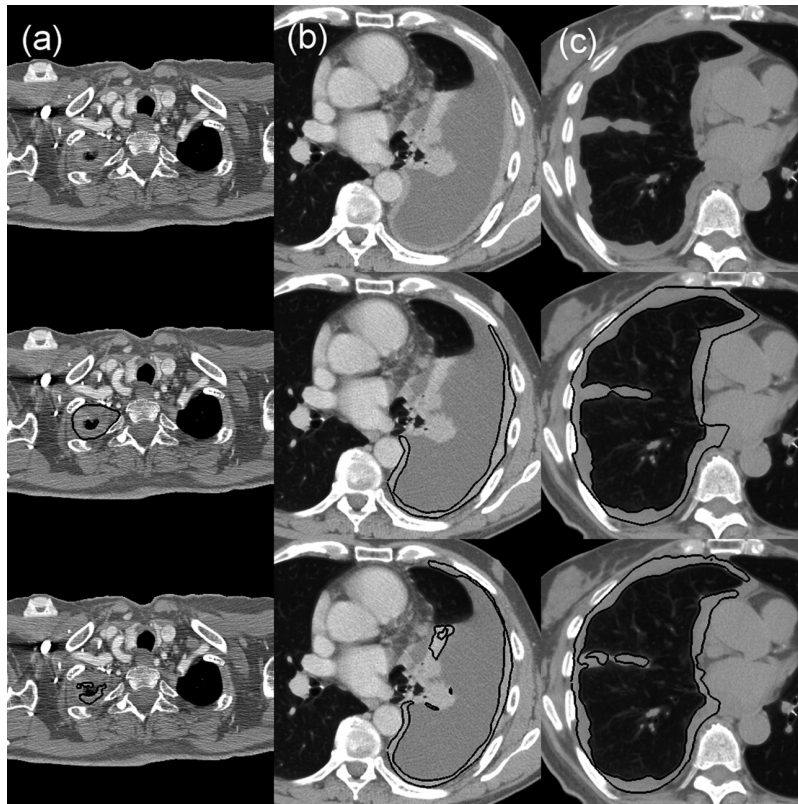


Fig. 10 Preprocessed CT sections (top), observer reference tumor segmentations (middle; black outlines), and predicted tumor segmentations by the present method (bottom; black outlines), for three sections from different CT scans of Test Set 2. Sections were selected at random from (a) the bottom 10th percentile (DSC = 0.374, AHD = 5.4 mm), (b) the interquartile range (DSC = 0.709, AHD = 13.0 mm), and (c) the top 10th percentile (DSC = 0.877, AHD = 0.8 mm) of the DSC values obtained when comparing predicted tumor segmentations of the present method and observer reference segmentations.

3.4 Interobserver Agreement for CT Sections that Exhibit Tumor and Effusion

Out of the 69 CT sections in the set of images used for the assessment of interobserver agreement, one or more of the five observers excluded pleural effusion from tumor contours on 26 sections (38%); the mean of the average interobserver DSC values on these 26 sections was 0.712 (median 0.743; range 0.512–0.853). The mean of the average interobserver DSC value on the 43 CT sections of this set for which none of the five observers excluded pleural effusion from tumor contours was 0.757 (median 0.779; range 0.517–0.915). The difference in the DSC distributions for these two subsets did not reach statistical significance using the two-sided Wilcoxon rank-sum test ($p = 0.07$).

4 Discussion

Mesothelioma patients commonly present with pleural effusion on imaging examinations; a majority of patients with this disease present with effusion at initial diagnosis.^{3,4} The robust, automated volumetric segmentation of mesothelioma tumor thus requires the proper differentiation of fluid from the adjacent tumor so that fluid is excluded from the segmented tumor volume. Our previous study on the deep learning-based segmentation of mesothelioma (“2018 Method”) showed a significantly improved segmentation performance when compared with a prior stepwise mesothelioma segmentation method; however, this deep learning-based method did not adequately exclude pleural effusion from tumor contours.^{22,23} Compared with

the 2018 Method, the present deep CNN-based mesothelioma segmentation method showed significantly greater overlap with radiologist-provided reference tumor contours on a test set of 94 CT sections (i.e., the “Tumor and Effusion Test Set”) that all exhibited both tumor and pleural effusion. The agreement between deep CNN-predicted tumor contours and observer-provided reference tumor contours on this test set, as evaluated using the AHD metric, was found to be significantly higher for the present method when compared with the 2018 Method. Bland–Altman plots comparing the segmented tumor area by the two deep learning-based methods with the observer-segmented tumor area on the Tumor and Effusion Test Set showed (1) a reduction in bias for the present method when compared with the 2018 Method and (2) a 95% CI for the mean relative bias of the present method that included 0. These results show a significant improvement in the performance of the present segmentation method when compared with the 2018 Method for the task of segmenting mesothelioma tumor on CT scans that exhibit pleural fluid.

The presence of pleural effusion on the CT scans of mesothelioma patients may increase observer variability in the task of mesothelioma segmentation due to the potentially unclear boundaries of tumor and fluid and the overlap in HU values between tumor and pleural fluid. A limitation of the present study was the inability to obtain interobserver comparisons on the Tumor and Effusion Test Set. To estimate the effect of pleural effusion on mesothelioma tumor contour variability among radiologists, the mean interobserver DSC value between pairwise combinations of five radiologists on a set of 69 axial CT sections was compared between (1) sections for which at least one radiologist excluded an area of pleural effusion from tumor contours and (2) sections for which none of the five radiologists excluded an area of pleural effusion from tumor contours. These sections were used in a previously published study on observer variability in mesothelioma tumor area measurements.³⁸ The difference in the DSC distributions for these two subsets did not reach statistical significance ($p = 0.07$); however, the lower mean interobserver DSC value for sections containing pleural effusion suggests that the concurrent presence of pleural fluid and mesothelioma tumor results in lower radiologist agreement in the task of mesothelioma tumor segmentation on CT scans.

Test Set 2 of this study included 130 CT sections from the two test sets that were used to evaluate segmentation performance in our previous study of the 2018 Method and provided a more general set of mesothelioma tumor presentations; 30% of the sections of Test Set 2 included pleural effusion. The present method did not show a significantly higher overlap with the set of observer-provided reference tumor contours on this test set when compared with the 2018 Method; however, the present method did achieve a significantly lower median AHD when compared with the 2018 Method on this test set. Bland–Altman analysis of the predicted tumor area by the two computerized methods on Test Set 2 showed a negative mean bias in predicted tumor area for the present method; however, the 95% limits of agreement for the relative difference in computerized tumor area and observer-segmented tumor area were narrower for the present method when compared with the 2018 Method. The improved agreement of the present method with radiologist-provided tumor contours on Test Set 2 shows that, despite the principal aim of the present method being the improvement of mesothelioma tumor segmentation on scans that exhibit both tumor and pleural effusion, the segmentation performance of the present method remains adequate across a test set for which the majority of cases do not exhibit pleural effusion.

Previous studies have found high interobserver variability in radiologist measurement of mesothelioma tumor.^{38,42} The median DSC value for radiologist interobserver comparisons was found to range from 0.65 to 0.81 across the two test sets of our previous study on the deep learning-based segmentation of mesothelioma.²³ Across both test sets of the present study, the overlap of deep learning-predicted tumor segmentations with radiologist tumor contours remained on par with radiologist interobserver overlap achieved on the two test sets of our previous study. Despite these encouraging results, the present method remains to be clinically validated through an observer study, whereby the segmentation performance of the method would be assessed by radiologists experienced in the measurement and assessment of mesothelioma tumor.

The present study trained deep CNNs separately for the segmentation of disease in the left and right hemithoraces. Across the 94 axial CT sections of the Tumor and Effusion Test Set,

there were no pixels erroneously predicted as tumor in the contralateral hemithorax. For the 130 sections of Test Set 2, there were three sections on which pixels of the contralateral hemithorax were erroneously included in the predicted segmentation by the present method. In one of these cases, 211 pixels (corresponding to a volume of 32 mm³) of the descending aorta and hilar vessels of the contralateral thorax were identified as tumor on a noncontrast-enhanced scan; in the other two cases, parts of a contralateral pleural effusion on two sections of the same CT scan were erroneously identified as tumor (71 and 833 pixels, corresponding to a volume of 24 and 282 mm³, respectively).

The method presented in this study did not incorporate three-dimensional (3D) context for the automated segmentation of mesothelioma tumor; the improvement in the segmentation performance when compared with the 2018 Method was achieved through the use of pretrained convolutional filters, an extensive validation methodology, and a more varied set of training sections that exhibited tumor with pleural effusion. Results on the validation sets of this study indicated that axial context could aid in the deep learning-based segmentation of mesothelioma on sections where the tumor was located adjacent to soft-tissue structures (e.g., medial tumor). The training of 3D CNNs requires higher-memory GPUs due to the larger image volumes that are processed during training and the increase in the number of parameters associated with 3D convolutional filters. Other deep CNN-based segmentation studies have employed image subsampling and/or downsampling to reduce the size of the image volumes used for training and testing.^{43,44} The variability in slice thickness across CT scans, combined with the anatomic extent and variability in appearance of mesothelioma tumor, precluded the development of a simple method for the subsampling of image volumes for this study. The use of downsampled lower-resolution CT volumes for training 3D CNNs was not pursued in this study due to the lack of fast, high-memory GPUs available for the training of the networks; furthermore, it is unclear to what extent gains in the segmentation performance achieved with increased axial context would overcome presumed reduction in the segmentation performance due to the lower resolution of the predicted tumor segmentations. An alternative to the full 3D approach is a more memory-efficient “2.5D” approach, which could allow for axial context to be incorporated in future studies for the full in-plane resolution segmentation of mesothelioma. For this approach, to provide additional 3D context, the network architecture would be modified to accept as input sections axially adjacent to the section for which tumor segmentation will be predicted; this technique has been applied to the task of liver segmentation.⁴⁵

5 Conclusions

This study implemented a deep learning-based method for the automated segmentation of mesothelioma tumor on CT scans, with the principal aim of improving the segmentation performance on scans of patients who presented with pleural effusion. Improvement in segmentation performance, when compared with our previously published study on deep learning-based segmentation of mesothelioma, was achieved through pretrained convolutional filters, an extensive validation methodology, and a more varied set of training sections that exhibited both tumor and pleural effusion. Significantly higher agreement with observer-provided tumor contours, in terms of segmentation overlap and the average distance between computerized and manual tumor contours, was found when compared with our previously published deep learning-based method on a test set of CT sections that exhibited both tumor and pleural fluid.

Disclosures

SGA receives royalties and licensing fees for computer-aided diagnostic technology through The University of Chicago.

Acknowledgments

The authors would like to acknowledge Anna K. Nowak, Jane E. Churpek, and Meghana Gadiraju for their assistance in the collection of mesothelioma patient scans used in the training

and test sets of this study. Partial funding for this work was provided by the NIH S10 RR021039 and P30 CA14599 grants. The contents of this paper are solely the responsibility of the authors and do not necessarily represent the official views of any of the supporting organizations. This work was supported, in part, by John D. Cooney and the firm of Cooney and Conway, and by the Plooy Family and the Kazan McClain Partners' Foundation, Inc.

References

1. N. J. Vogelzang et al., "Phase III study of pemetrexed in combination with cisplatin versus cisplatin alone in patients with malignant pleural mesothelioma," *J. Clin. Oncol.* **21**(14), 2636–2644 (2003).
2. G. Zalcman et al., "Bevacizumab for newly diagnosed pleural mesothelioma in the Mesothelioma Avastin Cisplatin Pemetrexed Study (MAPS): a randomised, controlled, open-label, phase 3 trial," *Lancet* **387**(10026), 1405–1414 (2016).
3. A. Kawashima and H. I. Libshitz, "Malignant pleural mesothelioma: CT manifestations in 50 cases," *Am. J. Roentgenol.* **155**(5), 965–969 (1990).
4. C. S. Ng, R. F. Munden, and H. I. Libshitz, "Malignant pleural mesothelioma: the spectrum of manifestations on CT in 70 cases," *Clin. Radiol.* **54**(7), 415–421 (1999).
5. A. C. Bibby et al., "Malignant pleural mesothelioma: an update on investigation, diagnosis and treatment," *Eur. Respir. Rev.* **25**(142), 472–486 (2016).
6. M. J. Byrne and A. K. Nowak, "Modified RECIST criteria for assessment of response in malignant pleural mesothelioma," *Ann. Oncol.* **15**, 257–260 (2004).
7. S. G. Armato, III and A. K. Nowak, "Revised modified response evaluation criteria in solid tumors for assessment of response in malignant pleural mesothelioma (version 1.1)," *J. Thorac. Oncol.* **13**(7), 1012–1021 (2018).
8. S. G. Armato, III et al., "Measurement of mesothelioma on thoracic CT scans: a comparison of manual and computer-assisted techniques," *Med. Phys.* **31**(5), 1105–1115 (2004).
9. G. R. Oxnard, S. G. Armato, III, and H. L. Kindler, "Modeling of mesothelioma growth demonstrates weaknesses of current response criteria," *Lung Cancer* **52**(2), 141–148 (2006).
10. H. I. Pass et al., "Preoperative tumor volume is associated with outcome in malignant pleural mesothelioma," *J. Thorac. Cardiovasc. Surg.* **115**(2), 310–318 (1998).
11. F. Liu et al., "Assessment of therapy responses and prediction of survival in malignant pleural mesothelioma through computer-aided volumetric measurement on computed tomography scans," *J. Thorac. Oncol.* **5**(6), 879–884 (2010).
12. T. Frauenfelder et al., "Volumetry: an alternative to assess therapy response for malignant pleural mesothelioma?" *Eur. Respir. J.* **38**(1), 162–168 (2011).
13. Z. E. Labby et al., "Disease volumes as a marker for patient response in malignant pleural mesothelioma," *Ann. Oncol.* **24**(4), 999–1005 (2013).
14. V. W. Rusch et al., "A multicenter study of volumetric computed tomography for staging malignant pleural mesothelioma," *Ann. Thorac. Surg.* **102**(4), 1059–1066 (2016).
15. H.-C. Shin et al., "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," *IEEE Trans. Med. Imaging* **35**(5), 1285–1298 (2016).
16. G. Litjens et al., "A survey on deep learning in medical image analysis," *Med. Image Anal.* **42**, 60–88 (2017).
17. O. Ronneberger, P. Fischer, and T. Brox, "U-net: convolutional networks for biomedical image segmentation," *Lect. Notes Comput. Sci.* **9351**, 234–241 (2015).
18. J. Deng et al., "ImageNet: a large-scale hierarchical image database," in *IEEE Conf. Comput. Vision and Pattern Recognit.*, pp. 248–255 (2009).
19. H. Ravishankar et al., "Understanding the mechanisms of deep transfer learning for medical images," *Lect. Notes Comput. Sci.* **10008**, 188–196 (2016).
20. N. Antropova, B. Q. Huynh, and M. L. Giger, "A deep feature fusion methodology for breast cancer diagnosis demonstrated on three imaging modality datasets," *Med. Phys.* **44**(10), 5162–5171 (2017).

21. B. Q. Huynh, H. Li, and M. L. Giger, "Digital mammographic tumor classification using transfer learning from deep convolutional neural networks," *J. Med. Imaging* **3**(3), 034501 (2016).
22. W. F. Sensakovic et al., "Computerized segmentation and measurement of malignant pleural mesothelioma," *Med. Phys.* **38**(1), 238–244 (2011).
23. E. Gudmundsson, C. M. Straus, and S. G. Armato III, "Deep convolutional neural networks for the automated segmentation of malignant pleural mesothelioma on computed tomography scans," *J. Med. Imaging* **5**(3), 034503 (2018).
24. N. Corson et al., "Characterization of mesothelioma and tissues present in contrast-enhanced thoracic CT scans," *Med. Phys.* **38**(2), 942–947 (2011).
25. A. K. Nowak et al., "The IASLC mesothelioma staging project: proposals for revisions of the T descriptors in the forthcoming eighth edition of the TNM classification for pleural mesothelioma," *J. Thorac. Oncol.* **11**(12), 2089–2099 (2016).
26. N. P. Campbell and H. L. Kindler, "Update on malignant pleural mesothelioma," *Semin. Respir. Crit. Care Med.* **32**(1), 102–110 (2011).
27. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv:1409.1556 (2014).
28. V. Iglovikov and A. Shvets, "TernausNet: U-net with VGG11 encoder pre-trained on ImageNet for image segmentation," arXiv:1801.05746 (2018).
29. E. Gudmundsson, C. M. Straus, and S. G. Armato III, "Pre-trained deep convolutional neural networks for the segmentation of malignant pleural mesothelioma tumor on CT scans," *Proc. SPIE* **10950**, 109503J (2019).
30. X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proc. 13th Int. Conf. Artif. Intell. and Stat.* (2010).
31. X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proc. 14th Int. Conf. Artif. Intell. and Stat.*, Vol. 15, pp. 315–323 (2011).
32. N. Srivastava et al., "Dropout: a simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.* **15**, 1929–1958 (2014).
33. D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," in *3rd Int. Conf. Learn. Represent.* (2014).
34. M. Abadi et al., "TensorFlow: a system for large-scale machine learning," in *12th USENIX Symp. Operat. Syst. Des. and Implementation*, pp. 265–284 (2016).
35. L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology* **26**(3), 297–302 (1945).
36. A. A. Taha and A. Hanbury, "Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool," *BMC Med. Imaging* **15**(1), 29 (2015).
37. M.-P. Dubuisson and A. K. Jain, "A modified Hausdorff distance for object matching," in *Proc. 12th Int. Conf. Pattern Recognit.*, IEEE, pp. 566–568 (1994).
38. Z. E. Labby et al., "Variability of tumor area measurements for response assessment in malignant pleural mesothelioma," *Med. Phys.* **40**(8), 081916 (2013).
39. S. Holm, "A simple sequentially rejective multiple test procedure," *Scand. J. Stat.* **6**(2), 65–70 (1979).
40. J. M. Bland and D. G. Altman, "Measuring agreement in method comparison studies," *Stat. Methods Med. Res.* **8**(2), 135–160 (1999).
41. J. M. Bland and D. G. Altman, "Statistical methods for assessing agreement between two methods of clinical measurement," *Lancet* **327**, 307–310 (1986).
42. S. G. Armato, III et al., "Observer variability in mesothelioma tumor thickness measurements," *J. Thorac. Oncol.* **9**(8), 1187–1194 (2014).
43. Ö. Cicek et al., "3D U-net: learning dense volumetric segmentation from sparse annotation," *Lect. Notes Comput. Sci.* **9901**, 424–432 (2016).
44. Q. Dou et al., "3D deeply supervised network for automated segmentation of volumetric medical images," *Med. Image Anal.* **41**, 40–54 (2017).
45. A. Ben-Cohen et al., "Fully convolutional network for liver segmentation and lesions detection," *Lect. Notes Comput. Sci.* **10008**, 77–85 (2016).

Eyjolfur Gudmundsson is a postdoctoral research scholar at the Centre for Medical Image Computing at University College London. He received his PhD in medical physics from

The University of Chicago in 2019. His thesis work with Dr. Samuel Armato at The University of Chicago was on the computer-aided diagnosis, segmentation and image analysis of malignant pleural mesothelioma on CT scans.

Christopher M. Straus is an associate professor of radiology and the director of medical student education at The University of Chicago. He received his AB and MD degrees from The University of Chicago in 1988 and 1992, respectively. He is the author of more than 70 journal papers and writing two book chapters. His current research interests center on optimizing medical education, imaging mesothelioma, and public perception of radiology.

Feng Li is a staff scientist in the Department of Radiology and a research radiologist of the Human Imaging Research Office at The University of Chicago. Her research interests include the detection of early lung cancers, analysis of radiologist-missed cancers, classification of malignant and benign lung nodules in chest CT scans or chest radiography, computer-aided diagnosis, observer performance studies, and advanced imaging techniques, such as machine learning applications and tumor response assessment.

Samuel G. Armato III is an associate professor of radiology and the Committee on Medical Physics at The University of Chicago. His research interests involve the development of computer-aided diagnostic (CAD) methods for thoracic imaging, including automated lung nodule detection and analysis in CT scans, semiautomated mesothelioma tumor response assessment, image-based techniques for the assessment of radiotherapy-induced normal tissue complications, and the automated detection of pathologic change in temporal subtraction images.