Auto Segmentation of Lung in Non-small Cell Lung Cancer Using Deep Convolution Neural Network

Auto Segmentation of Lung in Non-small Cell Lung Cancer Using Deep Convolution Neural Network

Ravindra Patil[1,2(✉)], Leonard Wee[1], and Andre Dekker[1]

[1] Department of Radiation Oncology (MAASTRO), GROW – School for Oncology and Developmental Biology, Maastricht University Medical Centre (MUMC), Maastricht, The Netherlands

[2] Philips Research Bangalore, Bengaluru, India

patil.ravindra@philips.com

Abstract. Segmentation of Lung is the vital first step in radiologic diagnosis of lung cancer. In this work, we present a deep learning based automated technique that overcomes various shortcomings of traditional lung segmentation and explores the role of adding "explainability" to deep learning models so that the trust can be built on these models. Our approach shows better generalization across different scanner settings, vendors and the slice thickness. In addition, there is no initialization of the seed point making it complete automated without manual intervention. The dice score of 0.98 is achieved for lung segmentation on an independent data set of non-small cell lung cancer.

Keywords: Lung segmentation · NSCLC · Deep learning

1 Introduction

Lung cancer is the leading cause for cancer related deaths and is accompanied by a dismal prognosis with a 5-year survival rate at only 18% [1]. Out of all Lung cancer, Non-Small Cell Lung Cancer (NSCLC) accounts for 85% of the cases. Treatment monitoring and analysis [2] using computed tomography (CT) images is an important strategy for early lung cancer diagnosis and survival time improvement. In these approaches, accurate Lung anatomy and pathology region segmentation is necessary as it directly related to the treatment plan. After decades of development in imaging techniques, volumes of high-resolution images with low distortions are now more easily available. Despite development of approaches for lung segmentation in recent years [3–6], achieving accurate segmentation performance continues to require attention because of specific challenges. One such example is tumors have an intensity similar to that of lung wall; thus, they are difficult to distinguish using intensity values alone and also the structure of the lung changes based on the disease pathology such as consolidation, masses, pneumothorax or effusions.

There is considerable progress in developing Lung segmentation algorithms that have an ability to perform accurate delineation under different disease conditions [7].

M. Singh et al. (Eds.): ICACDS 2020, CCIS 1244, pp. 340–351, 2020.
https://doi.org/10.1007/978-981-15-6634-9_31

In general, all the lung segmentation algorithms can be classified into following five sub categories (1) Intensity based (2) Shape based or Model based (3) Neighboring anatomy guided (4) Region based (5) Artificial Intelligence based [7]. The intensity based approaches are fast, intuitive and computationally efficient however these techniques fail during the pathological condition where there are attenuation variation. The shape based or the model based approaches provide a very good accuracy due to template mapping, however these algorithms are computationally inefficient and it's difficult to create representative training features. Neighboring anatomy based approach exploits the information of the spatial context of the neighboring organ of the lung such as rib cage, heart, spine for extracting the contours of the lung region, this approach is computationally expensive but provide good results when the intensity variation is mild to moderate. However, in case of the extreme diseased condition such as opacification of entire hemithorax this approach fails. Region growing approaches such as watershed transform, graph cuts and random walks are efficient but they tend to over segment. In recent times AI based (Machine Leaning and Deep Learning) approaches have become popular and in particular Deep Learning (DL) approaches due to the better accuracy that these algorithms achieve in ill defined pathologic conditions [8].

Recently, Convolutional Neural Networks (CNN) have been seen as a powerful tool for learning features from network layers [9]. CNN's act as a tool for learning discriminative features, which are useful in different image processing and computer vision tasks. CNN's need relatively less pre-processing compared to other known algorithms, which means the network learns the filters that in traditional algorithms had to be hand engineered [9]. The independence from prior knowledge and effort required in feature selection design is a major advantage. Moreover, CNN's include multi-layer processing, which ensures that the model learns the features at the granular level. In recent time, there has been progressive usage of CNN's for various medical segmentation tasks. Dou et al. [10] uses a 3D CNN, which focuses on the task of automatic nodule detection. DIAG Convnet by Setio et al. [11] provides automatic pulmonary nodule detection in CT images. The review article by Zhou et al. [12] lists various deep learning based medical segmentation approaches developed for different modalities and anatomies. Although, CNN's are explored for medical image segmentation, we did not come across a study with large-scale validation of the Lung segmentation approach with deep learning in diseased conditions.

The objectives of the current work are manifold. The first being to test the efficacy of the deep learning models on a large scale diseased Lung dataset to automatically segment the Lung region. Second, to compare DL approach with the traditional segmentation methods and comment on the complexity and efficiency. Third being, adding "explainability" to deep learning models so that the trust can be built on these models.

2 Methodology

In the traditional approach of Lung segmentation, the canonical steps are employed as shown in Fig. 1. However, there might be slight variation in each of the block based on the algorithm that is considered for implementation. Each of these blocks needs tuning based on the image type and the acquisition parameters of the scanner. Marker generation

is used to define the region that is present inside and outside the Lung region; this is performed manually by marking the region using a seed point. Further, the image is preprocessed by applying the filters as well as HU thresholding to eliminate unwanted region. This is followed by running a segmenter algorithm (such as watershed, active contour etc.) and performing the morphological and post processing operations to correct the contours. These traditional approaches needs manual and empirical tuning and quite difficult to generalize on the large and varied dataset with different acquisitions and threshold values.

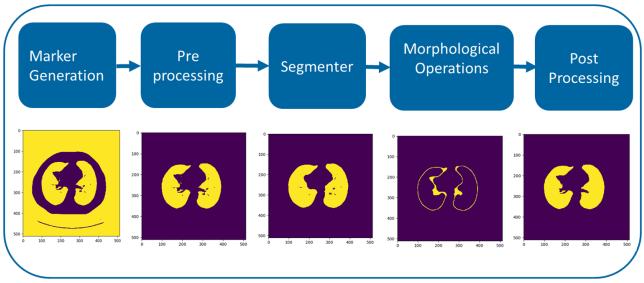


Fig. 1. Traditional lung segmentation approach

The approach adapted in this work is described in the section below, which is represented by the block diagram schematic as shown in Fig. 2. There are two phases in creation of the deep learning model, which aids in segmentation of the Lung anatomy. First being the training phase, where in the data is fed after preprocessing to the model, where the model is trained with the annotated ground truth (region of the Lung) as the reference. Further, the learnt model layers are analyzed using visualization to ascertain, what region of the image, model looked into to arrive at the delineation of the Lung region. In the scoring phase, the trained model is used to delineate the region of the lung on the unseen/live Lung CT scans.

2.1 Data

The data set used for model training which was obtained from The Cancer Imaging Archive (TCIA) repository of NSCLC patients [13]. This dataset contained pretreatment CT scans where in the lung regions were manually delineated by the radiation oncologist on the 3D volume. This data set will be referred as Lung 1. In total 422 subjects were used for training, which maps to ~42,019 Lung CT slices on which the algorithm was trained. Out of 422 subjects, randomly 300 subjects were used for model training and remaining 122 for model validation. The training data had following sub categories of NSCLC: Adenocarcinoma, Large cell carcinoma and Squamous cell carcinoma. The CT scanners used for imaging the subjects were from different vendors (Siemens, CNS Inc, Philips and GE) and the slice thickness varied from 0.65 to 5 mm with 512 * 512 resolution. Further, for testing the model, independent data set of Lung 2 was used which is marked as NSCLC Radiogenomics dataset in TCIA repository with 211 subject's [14].

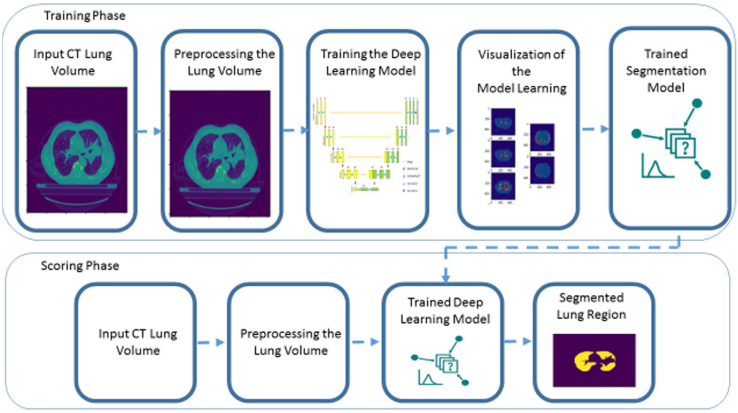
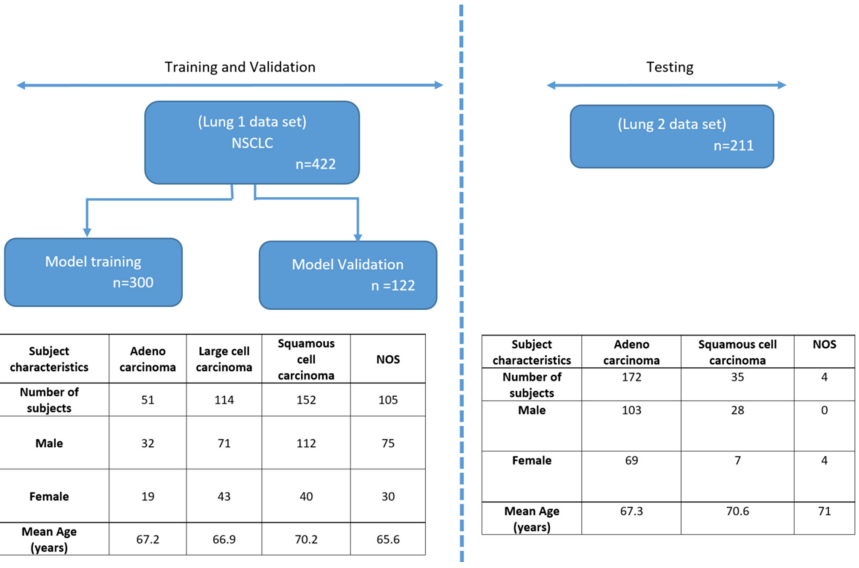


Fig. 2. Deep Learning based lung segmentation approach

The testing data had following different sub categories of NSCLC: Adenocarcinoma, Squamous cell carcinoma and NOS. The demographics of each of the subgroups of training and testing are mentioned in the Fig. 3.



Subject characteristics	Adeno carcinoma	Large cell carcinoma	Squamous cell carcinoma	NOS
Number of subjects	51	114	152	105
Male	32	71	112	75
Female	19	43	40	30
Mean Age (years)	67.2	66.9	70.2	65.6

Subject characteristics	Adeno carcinoma	Squamous cell carcinoma	NOS
Number of subjects	172	35	4
Male	103	28	0
Female	69	7	4
Mean Age (years)	67.3	70.6	71

Fig. 3. Data demographics and data split for model training

2.2 Pre Processing

The input CT volumes are processed to set non-anatomical regions such as air (with HU value below −1000) to 0, so that number of the pixel computations are reduced.

The data augmentation was performed in terms of translation and rotation (0 to 30°) to make the model robust against the data variations. No specific preprocessing steps such as de-noising, artifact corrections, system-based calibration were employed, which are typical in conventional approaches of Lung segmentation. The preprocessing approach was specifically designed to have minimum steps, to test the efficacy of convolution neural networks.

2.3 Training Deep Learning Model

The target of the current approach is, given a CT slice of a lung (diseased or non-diseased) the region of whole Lung need to be segmented. We employ modified U-Net [15] inspired CNN architecture to arrive at segmented region of the lung. We follow a pixel based classification mechanism that aims at classifying whether each pixel belongs to Lung region. The model architecture that is built was inspired from U-Net Convolution Neural Network (CNN), which consists of 18 convolutional layers, 4 central pooling layers with the convolutional kernel size of 3X3 in each convolutional layer. The schematic representation of the model architecture is as shown in the Fig. 4.

The convolutional layers perform convolutional operation on all input feature maps to obtain output features defined by the Rectified Linear Units (ReLU) activation function [16]. The feature map layer combination is defined by the Eq. (1)

$$f^j = ReLU\left(\sum_{i=1}^{m} C^{ij} * f^i + b^j\right) \tag{1}$$

Where f^i and f^j are the i^{th} input feature map and j^{th} output feature map, respectively. We define C^{ij} as the convolutional kernel between f^i and f^j (* denotes the 2-D convolutional operation), b^j is the bias of the j^{th} output feature map.

After each convolutional layer, a rectified linear unit (ReLU) is used as a non-linear activation function, this is added to bring non linearity to the model and is expressed as:

$$ReLU(z) = \max(0, z) \tag{2}$$

Further to the last convolutional layer, a fully connected layer is applied where each output unit connects to all inputs. This layer can capture correlations between different features produced by the convolutional layer. For achieving non-linearity and a two-class output classifier, the sigmoid function was used. Since the sigmoid function ranges from zero to one, it can be directly related to class probabilities making it ideal activation function for classification task.

$$sigmoid(z) = \frac{1}{1 + e^{-z}} \tag{3}$$

The goal of network training is to maximize the probability of the correct class. This is achieved by minimizing the dice coefficient loss function. The loss function is minimized during the model's training process. The weight updation was performed using Adaptive Moment Estimation (ADAM) algorithm [17]. Instead of adapting the parameter learning rates based on the average first moment as in RMSProp, ADAM makes use of the average of second moments of the gradients. Specifically, the algorithm

calculates an exponential moving average of the gradient and the squared gradient, and the parameters β_1, β_2 control the decay rates of these moving averages. The initial value of the moving averages and β_1, β_2 values close to 1.0 (recommended) results in a bias of moment estimates towards zero. This bias is overcome by calculating the biased estimates and then calculating bias-corrected estimates. ADAM is an extension to stochastic gradient descent and converges faster than other stochastic optimization methods [18]. It also rectifies problem such as vanishing learning rate, slow convergence that other optimization problems face which leads to fluctuating loss function. The weights are updated based on the below equation

$$w^{(t+1)} \leftarrow w^{(t)} - \eta \frac{\widehat{m_w}}{\sqrt{\widehat{v_w}} + \varepsilon} \tag{4}$$

Where ε is a small number used to prevent division by zero.
And

$$\widehat{m_w} = \frac{m_w^{(t+1)}}{1 - \beta_1^t} \tag{5}$$

$$\widehat{v_w} = \frac{v_w^{(t+1)}}{1 - \beta_2^t} \tag{6}$$

Where m_w and v_w are estimates of the first moment and second moment of the gradients respectively. β_1 and β_2 are the forgetting factors for gradients and second moments of gradients, respectively.

The hyperparameters used for the model training are Optimizer = ADAM, Learning Rate = 1.0e−6, Metric = Dice score, Number of Epochs = 50, Batch Size = 2, Weight Initialization Method = Xavier initialization. The dice similarity coefficient (DSC) is used as the primary evaluation criteria for assessing the automatic segmentation accuracy; also, this is used as a loss function for the backward propagation in the proposed model. The DSC expressed as in Eq. (7) provides amount of overlap between two segmentation results [19], wherein G_t is the groundtruth segmentation and *Auto* is the automated segmentation performed by the trained model.

$$DSC = \frac{2 * V(G_t \cap Auto)}{V(G_t) + V(Auto)} \tag{7}$$

In the implementation, RT Structure of delineated lung region on each slice by the radiologist is considered as the ground truth and the region predicted by the model is overlaid on the RT Structure mask to arrive at the DSC.

2.4 Model Visualization

Interpreting the deep learning model and understanding model's rationale behind the decision process to arrive at the prediction is challenging. In order to ensure that the model is indeed looking into the relevant regions in the input image to arrive at the decision, a visualization engine was added on top of the built model. The approach used to

build the model visualization was Gradient-weighted Class Activation Mapping (Grad-CAM) [20]. The Grad-CAM works on the class discriminative localization approach wherein it uses the gradients of target of the final convolutional layer to produce a coarse localization map, highlighting the regions considered in the image for predicting the concept. This is not only useful to know about the regions responsible for prediction but also aids in debugging the decision process in the networks. In essence, Grad-CAM takes into account the penultimate layer (layer before softmax) to interpret the decision of the CNN and identifies the respective filter activation for every spatial location (i, j) in the given image, further this is converted into the heat map based on the weights indicating the prominent regions. The mathematical aspect is depicted in the Eq. (8, 9).

$$W_k^c = \frac{1}{z} \sum_i \sum_j \frac{\partial Y^c}{\partial A_{ij}^k} \tag{8}$$

$$S^c = \frac{1}{z} \sum_i \sum_j \sum_k w_k^c A_{ij}^k \tag{9}$$

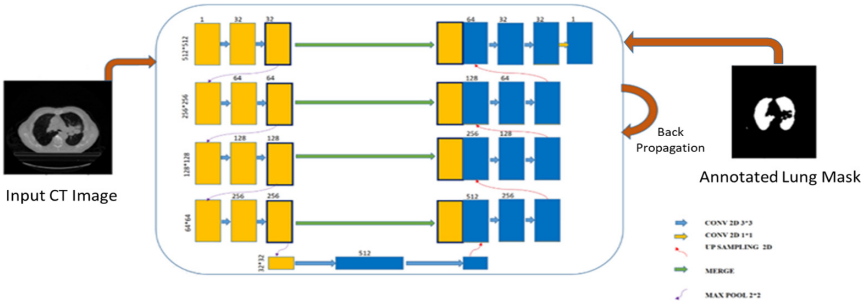


Fig. 4. Model architecture

The spatial score of a specific class S^c is the global average pooling over spatial location (i, j) for the gradient of respective class output Y^c with respect to the feature map A_{ij}^k. The spatial score is obtained by multiply the resulting value with the feature map along with its channel axis k. The $\sum$ describes the pooling and average operation and z is constant. The output of the sample GradCAM results can be seen in Fig. 5. It can be observed from the figure that location of the lungs are highlighted with red colour indicating the maximum activation of the filters in that region, mapping to understanding that the model has learnt the region of image that needs to be segmented.

3 Results and Discussion

The training data set containing 300 subjects of NSCLC, were fed to the U-Net inspired model for training and 122 subjects were used as the validation set. The training accuracy of 0.99 dice score was obtained after 50 epochs of training. The quantitative analysis resulted in average dice score of 0.98 on the independent test data of Lung 2 dataset,

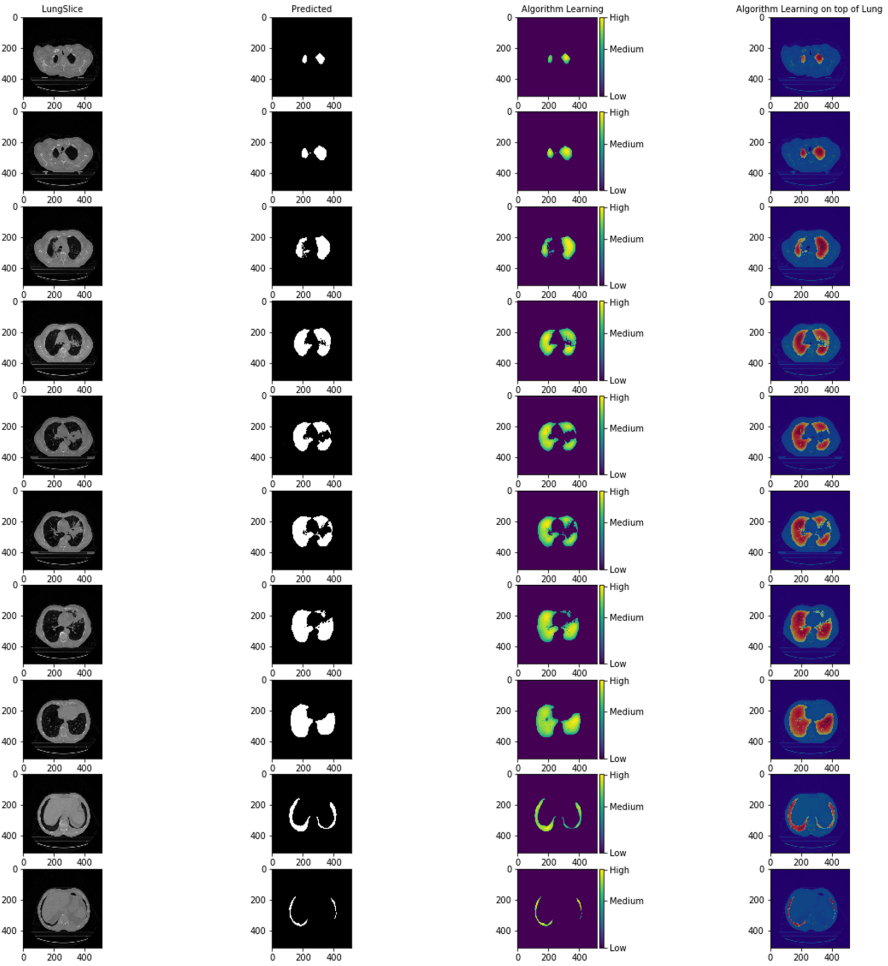


Fig. 5. GradCAM visualization on Lung region.

comprising of 211 subjects. The Bland-Altman plot showing the variation of difference between the ground truth dice score and predicted dice score vs mean of ground truth and predicted dice score of the test data is plotted in the Fig. 6. It can be observed that the differences are within mean ± 1.96 SD indicating that the segmentation performed by the model resembles the ground truth with high accuracy. Also, the box plot in the Fig. 7 on the test data shows the similar results indicating the average predicted dice score in the range of 1 to 0.98 without outliers being seen, thus suggesting the robustness of the model. The model was built using Keras (2.1.1) and Tensor flow (1.2.1) as backend in python 3.0. The machine configuration used for training the model had Nvidia GPU (Titan X 1080Ti) with the mentioned model hyper parameters. The scoring time on the same GPU configuration took one millisecond per lung slice.

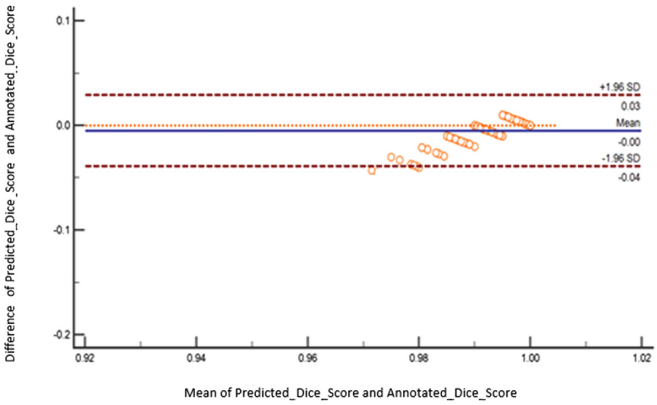


Fig. 6. Bland-Altman plot of dice score of lung anatomy

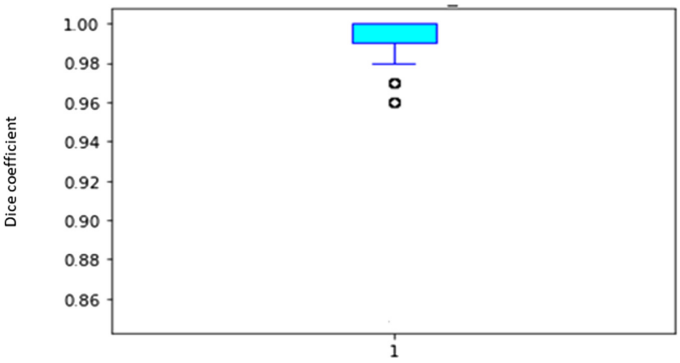


Fig. 7. Box Plot of dice score of lung anatomy on the test set

The sample results of segmented output can be observed in the Fig. 8. The first column is input CT slice, second column depicts annotation of the lung marked by the radiation oncologist, third column in the figure maps to the segmented output by our model, fourth being lung contour extracted from the predicted segmentation and fifth column is the over laid contour on the original Lung slice. It can be observed that the lung segmentation model is robust to the different anatomical structure of the Lung ranging from symmetric lung (rows d–f) to partial lung region being visible (row i–j), in the bottom two rows of the figure.

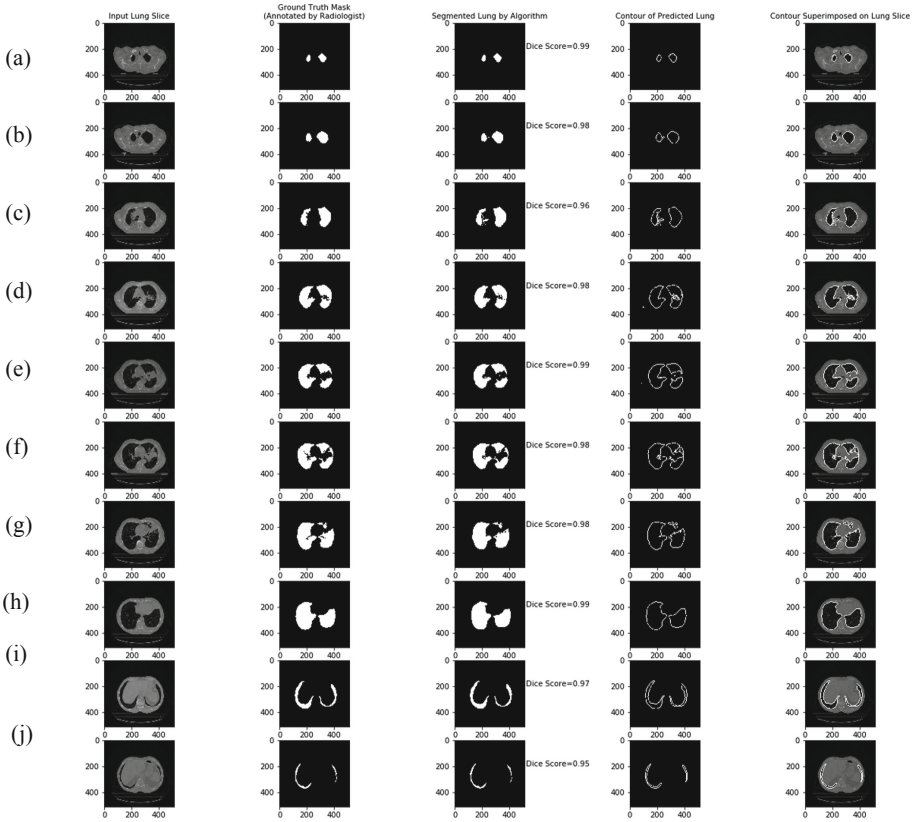


Fig. 8. Segmented Lung region and the associated ground truth masks

Table 1. Comparison with other lung segmentation approaches.

Title	Approach	Number of dataset(n)	Average dice score	Test dataset
Y. Wei G et al. [21]	Bresenham algorithm	97 subjects	0.95	Externally validated on 25 Subjects
Dai et al. [22]	Gaussian Mixture based model	Not mentioned	0.98	No external validation
Noor et al. [5]	Thresholding & morphological	96 subjects	0.98	No external validation
Zhang et al. [23]	Active contour	60 subjects	0.97	Externally validated on 60 subjects
Our approach	Deep Learning	422 Subjects	0.98	Externally validated on 211 subjects

4 Conclusion

Multiple algorithms in the literature have attempted the segmentation of the lung anatomy, the algorithms varies from simple thresholding to that of machine learning approaches. The Table 1 provides comparison of lung segmentation approaches being published recently. The prior work by Y. Wei et al. [21] performs both preprocessing as well as post processing of the scans to arrive at the segmentation and its validated on limited external data of 25 subjects leading to DSC of 0.95. However, as there are preprocessing steps used in this study, where in each of these steps needs to optimize based on the scan type, slice thickness and acquisition parameters leading to loss of generalization in the approach. Another study by Dai et al. [22] used Gaussian mixture model based approach; wherein there is need of manual seed point, initialization and custom preprocessing for noise elimination. The study does not mention the data characteristics, diseased conditions as well as number of subjects used for building the model, leading to questioning the claims of the study. Noor et al. [5] used thresholding approach and claimed DSC of 0.98 without external validation data set. In addition, the study focused only on a single scanner type as well as single slice thickness of 10 mm making the approach difficult to be generalized across all the scanner types, disease conditions and the variability. Another study by Zhang et al. [23] showed robustness of active contour model for delineation of the lung region with pathological conditions, also the model took into account the variability of scanner setting and the data. Further, the model was validated on the external data of 60 subjects resulting in the dice score of 0.97, with a robust validation approach. However, in this study there is a need to initialize the threshold value and perform preprocessing steps such as Gaussian smoothing. Compared to existing approaches, our model is exhaustively validated on the diseased data set with external data and there is no preprocessing or post processing performed, no domain-based adaptation is performed to arrive at the delineation. Our approach shows better generalization across different scanner settings, vendors and the slice thickness. In addition, there is no initialization of the seed point making it complete automated without manual intervention.

References

1. Siegel, R., Miller, K., Jemal, A.: Cancer statistics, 2016. CA-Cancer J. Clin. **66**, 7–30 (2016)
2. Aerts, H., et al.: Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat. Commun. **5**, 4006 (2014). https://doi.org/10.1038/ncomms5006
3. Wei, Y., Shen, G., Li, J.J.: A fully automatic method for lung parenchyma segmentation and repairing. J. Digit. Imaging **26**(3), 483–495 (2013)
4. Dai, S., Lu, K., Dong, J., Zhang, Y., Chen, Y.: A novel approach of lung segmentation on chest CT images using graph cuts. Neurocomputing **168**, 799–807 (2015)
5. Noor, N.M., et al.: Automatic lung segmentation using control feedback system: morphology and texture paradigm. J. Med. Syst. **39**(3), 22 (2015)
6. Pulagam, A.R., Kande, G.B., Ede, V.K.R., Inampudi, R.B.: Automated lung segmentation from HRCT scans with diffuse parenchymal lung diseases. J. Digit. Imaging **29**(4), 507–519 (2016). https://doi.org/10.1007/s10278-016-9875-z

7. Awais, M. et al.: Segmentation and image analysis of abnormal lungs at CT: current approaches, challenges, and future trends. Radiographics (2015). https://doi.org/10.1148/rg.2015140232
8. Chae, S.H., Moon, H.M., Chung, Y., Shin, J., Pan, S.B.: Automatic lung segmentation for large-scale medical image management. Multimed. Tools Appl. **75**(23), 15347–15363 (2016)
9. Yamashita, R., Nishio, M., Do, R.K.G., et al.: Convolutional neural networks: an overview and application in radiology. Insights Imaging **9**, 611–629 (2018). https://doi.org/10.1007/s13244-018-0639-
10. Dou, Q., Chen, H., Yu, L., Qin, J., Heng, P.A.: Multi-level contextual 3D CNNs for false positive reduction in pulmonary nodule detection (2016). IEEE Trans. Biomed. Eng. https://doi.org/10.1109/tbme.2016.2613502
11. Setio, A.A.A.: Pulmonary nodule detection in CT images: false positive reduction using multi-view convolutional networks. Med. Imaging IEEE Trans. https://doi.org/10.1109/TMI.2016.2536809
12. Zhou, T., et. al.: A review: deep learning for medical image segmentation using multi-modality fusion. Array **3**, 10004 (2019). https://doi.org/10.1016/j.array.2019.100004
13. Aerts, H.J., et al.: Data from NSCLC-radiomics [data set]. Cancer Imaging Arch (2019). https://doi.org/10.7937/K9/TCIA.2015.PF0M9REI
14. Shaimaa, B., et al.: Data for NSCLC radiogenomics collection. Cancer Imaging Arch. (2017). http://doi.org/10.7937/K9/TCIA.2017.7hs46erv
15. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, William M., Frangi, Alejandro F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
16. Xu, B., Wang, N., Chen, T., Li, M.: Empirical evaluation of rectified activations in convolutional network. arXiv preprint arXiv:1505.00853 (2015)
17. Goodfellow, I., Bengio, Y., Courville, A., Bengio, Y.: Deep Learning, vol. 1. MIT press, Cambridge (2016)
18. Ruder, S.: An overview of gradient descent optimization algorithms. arXiv preprint arXiv:1609.04747 (2016)
19. Havaei, M., Guizard, N., Chapados, N., Bengio, Y.: HeMIS: hetero-modal image segmentation. In: Ourselin, S., Joskowicz, L., Sabuncu, Mert R., Unal, G., Wells, W. (eds.) MICCAI 2016. LNCS, vol. 9901, pp. 469–477. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46723-8_54
20. Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. arXiv:1610.02391 (2016)
21. Wei, Y., Shen, G., Juan-juan, L.: A fully automatic method for lung parenchyma segmentation and repairing. J. Digit. Imaging **26**, 483–495 (2013). https://doi.org/10.1007/s10278-012-9528-9
22. Dai, S., Lu, K., Dong, J., Zhang, Y., Chen, Y.: A novel approach of lung segmentation on chest CT images using graph cuts. Neurocomput. **168**(C), 799–807 (2015). https://doi.org/10.1016/j.neucom.2015.05.044
23. Zhang, W., Wang, X., Zhang, P., Chen, J.: Global optimal hybrid geometric active contour for automated lung segmentation on CT images. Comput. Biol. Med. **91**, 168–180 (2017). https://doi.org/10.1016/j.compbiomed.2017.10.005