An End-to-End Deep Learning Pipeline for Emphysema Quantification Using Multi-label Learning

Mohammadreza Negahdar, Adam Coy, and David Beymer

Abstract— **We propose and validate an end-to-end deep learning pipeline employing multi-label learning as a tool for creating differential diagnoses of lung pathology as well as quantifying the extent and distribution of emphysema in chest CT images. The proposed pipeline first employs deep learning based volumetric lung segmentation using a 3D CNN to extract the entire lung out of CT images. Then, a multi-label learning model is exploited for the classification creation differential diagnoses for emphysema and then used to correlate with the emphysema diagnosed by radiologists. The five lung tissue patterns which are involved in most lung disease differential diagnoses were classified as: ground glass, fibrosis, micronodules (random, perilymphatic and centrilobular lung nodules), normal appearing lung, and emphysematous lung tissue. To the best of our knowledge, this is the first end-to-end deep learning pipeline for the creation of differential diagnoses for lung disease and the quantification of emphysema. A comparative analysis shows the performance of the proposed pipeline on two publicly available datasets.**

I. Introduction

Emphysema is a progressive disease due to permanent enlargement of air spaces distal to terminal bronchioles accompanied by destruction of alveolar walls. Volumetric thoracic CT imaging allows the study of these morphological changes, whereupon diagnosis and quantification of emphysema phenotype [1, 2]. Assessing the volume of the emphysematous tissue is needed to evaluate disease severity and progression, and to predict future risks such as lung cancer [3-5].

To assist clinicians and improve their diagnostic performance, computer aided diagnosis systems (CAD) have been extensively developed. The imaging data are interpreted through the assessment of the existence and extent of emphysema as well as various textural patterns within the lung [6, 7]. Although both restrictive and obstructive lung diseases are histologically heterogeneous group of diseases, they share a similar clinical manifestation with each other called chronic obstructive pulmonary disease (COPD). Thus, the diagnosis of emphysema cannot be accurately addressed as a single-label texture analysis problem and scarcity of precisely annotated data does not allow for this problem to be addressed as a segmentation problem.

Since restrictive and obstructive lung diseases are manifested as textural alterations in the lung parenchyma, convolutional neural networks (CNN) and deep learning models show tremendous potential for learning features to classify different lung tissues [8-11]. However, most studies have tackled the emphysema detection problem as a single-label classification problem [10, 12].

In this study, we propose an end-to-end pipeline based on CNN for thoracic CT images to accurately evaluate the existence and proportion of emphysematous lung tissue. We have already developed a volumetric deep learning-based lung segmentation algorithm using 3D CNN to accurately segment the lung from CT images [13]. In this paper, we build an infrastructure based on multi-label deep learning model to differentiate and classify lung tissues within the segmented lung to detect emphysema in the presence of the co-occurring lung diseases. With respect to tissue labels, this multi-label learning bears direct resemblance with a radiologist's diagnostic through process.

We evaluate the performance of the proposed pipeline using two publicly available datasets. We report the classification performance and emphysema quantification results on both the patch-level and patient-level. We reserve one dataset solely for validation and performance evaluation of the proposed pipeline.

II. Material and Method

Dataset: We used a publicly available multimedia dataset of interstitial lung diseases from the Geneva University Hospital (HUG) [14]. This dataset consists of 109 high resolution CT of different interstitial lung disease (ILD) cases with 512x512 pixels per slice and an average of 25 slices per image. The slice thickness is 1-2 mm and slice spacing is 10-15 mm. Manual annotation for 17 different lung patterns along with clinical parameters from patients with histologically proven diagnoses of ILDs are provided. Since annotations do not cover the entire lung, we reserved this dataset for training and we did not use it for validation purposes. Therefore, we used the lung tissue research consortium (LTRC) dataset, which provides a complete labeling of the entire lung [15]. The LTRC dataset consists of patients of different restrictive and obstructive lung diseases with 512x512 pixels per slice.

In order to establish an accurate and scalable end-to-end pipeline of emphysema quantification, we chose the five lung tissue patterns involved in the most restrictive and obstructive differential lung disease diagnoses: ground glass, fibrosis, micronodules, normal appearance, and emphysema. Therefore, we can reliably differentiate and diagnose emphysema against normal lung tissue as well as other diseases which can be misdiagnosed with emphysema. Fig. 1 shows the flowchart of the proposed end-to-end pipeline.

Mohammadreza Negahdar is with IBM Research – Almaden, San Jose, CA 95120 USA (phone: 408-927-1282; e-mail: mnegahd@us.ibm.com).

Adam Coy is with the Department of Radiology and Biomedical Imaging, University of California San Francisco (UCSF), San Francisco, CA, USA (e-mail: acoy@ us.ibm.com).

David Beymer is with IBM Research – Almaden, San Jose, CA 95120 USA (e-mail: beymer@us.ibm.com).

978-1-5386-1311-5/19/$31.00 ©2019 IEEE

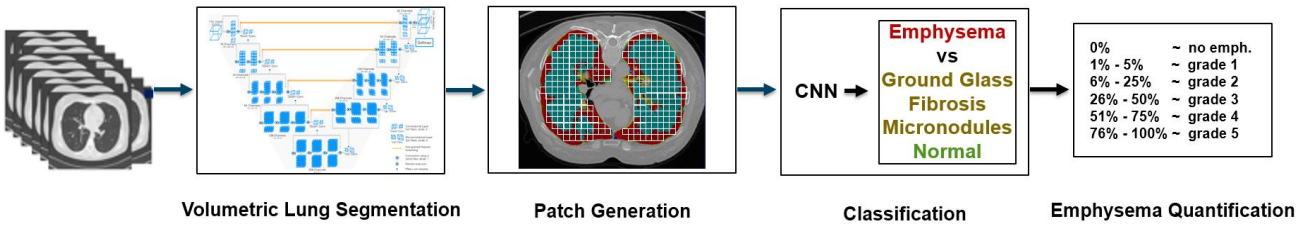


Figure 1: Proposed end-to-end pipeline for Emphysema quantification.

A. *Lung Segmentation*

We employed a model based on a modified V-net [16], which exploits both local features as well as global context to volumetrically segment the lung [13]. Thereby, we assured that various slice spacing would not affect the accuracy of lung segmentation; and hence emphysema quantification. Since our field of interest is limited to lung parenchyma, we detected and excluded the bronchovascular tree from segmented lung.

B. *Patch generation*

Each class of the aforementioned five lung tissue textures is a stochastic repetition of pulmonary pathology. Thus each lung tissue pattern is characterized by local features which leads us toward local (patch) classification. Therefore, each slice of lung is divided into patches of 32x32 pixels which has complete overlap with the segmented lung mask.

For training, each generated patch has 100% overlap with the segmented lung and at least 80% overlap with one of five lung tissue classes in an annotated image. To alleviate the HU variability due to different CT scanners and reconstruction kernels, the image intensity value of each patch is cropped within [-1000, 200] HU, and then mapped to [0, 1].

C. *Classification Architecture*

We have already evaluated and compared four different models based on four mainstream network architectures for lung tissue characterization [17]: AlexNet [18], GoogLeNet [19], ResNet [20] and Densenet [21]. To achieve a robust and fast network with higher specificity, in this study, we employed a network architecture based on AlexNet by replacing the softmax single-label loss with a multi-label loss function. This modified loss function allows multi-label learning, and consequently lung tissue differentiation.

CNNs highlight small repetitive structures that resemble the convolution kernel, so the size and number of employed CNNs in our applied model affects specificity and sensitivity of the learned model [17].

We generated patches of 32x32 pixels in order to obtain relevant local structures and minimize irrelevancy. For classification, we up-sampled our patches to a size of 227x227 pixels to allow training a deep CNN with multiple layers of down-sampling. Our network consists of a series of five CNN layers: 11x11, 5x5, 3x3, 3x3, and 3x3. We obtained 4096 feature maps in the last layer which was passes through a series of 3 dense layers. The final output was a series of five numbers which showed the probability of the patch classes. We employed a step function whose binarizing threshold is defined on 0.8 in order to determine the presence or absence of emphysema.

D. *Training and testing*

For training, we generated 18793 non-overlapping 32x32 pixels patches within the segmented lung from the HUG dataset. We equalized sample distributions across the classes. Since emphysema is the main focus of this study, we generated emphysema patches of patients with different patterns of emphysema (centrilobular, paraseptal, and panlobular). To increase the number of training patches and prevent over-fitting, these patches were augmented by applying random affine transformations. We randomly picked 80% of all patches for training and reserved the rest for validation. We evaluated the performance of our model on the LTRC dataset. We adopted the Adam optimizer to minimize a categorical cross entropy loss function. We also used a fixed learning rate of 0.0001 and trained the model for 500 epochs with a batch size of 128 using Keras with a TensorFlow backend.

E. *Emphysema quantification*

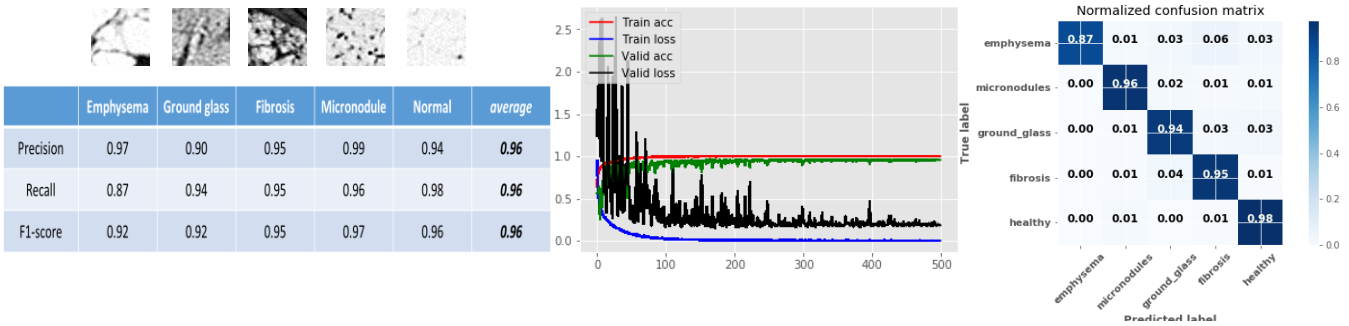
For the patient-level study, each patient is assigned a categorical grade based on the percentage of emphysematous patches in the lung. Patients are graded from 1 to 5 corresponding to 1%-5%, 6%-25%, 26%-50%, 51%-75%, and 76%-100% emphysematous lung volume, respectively [22]. In case of finding no emphysema, patient is graded 0.

III. RESULTS

Fig. 2 shows the learning curve and confusion matrix of our proposed multi-label classifier. Validation loss consistently decreases and both training and validation accuracy show notably fast convergence. Our training scheme and data augmentation allows us to train the proposed model on a reasonable sized dataset without overfitting.

For performance evaluation of the proposed classifier on the validation data, as the ubiquitous evaluation measures, we report precision, recall, and f1-scores for all 5 classes; as presented in Fig. 2.

To assess the overall performance of the proposed end-to-end pipeline, 11 representative patients of LTRC dataset who have different types of emphysema (centrilobular, paraseptal, and panlobular) have been selected. Performance of the proposed pipeline in both patch-level and patient-level scales are presented in Table I. Emphysema presence has been



(a) (b) (c)

Figure 2: (a) Precision, recall, and f1-score of the proposed classifier. Examples of each lung tissue patches in training set is shown above each associated column. (b) Learning curve for training and (c) calculated normalized confusion matrix of multi-label classifier for generated patches.

accurately predicted in all patients. In addition, emphysema grades are correctly estimated in 9 patients out of the 11 reported patients in this study. For the other two patients, the estimated grades with the proposed pipeline are distanced only 1 step from the annotated dataset.

Next, 1056 patients from the LTRC dataset have been randomly selected and evaluated. In order to perform emphysema differential diagnosis, the percentage of emphysematous patches in the lung is used as an approximation to the probability of the patient having emphysema. Given patient-level ground-truth labels of emphysema, the best cut-off probability can be estimated for a given target sensitivity. Then for that target, predicted emphysema labels are compared against emphysema diagnosed by radiologists. Evaluation has been performed based on the calculated specificity of diagnosed emphysema patients at the 20% and 50% sensitivity levels, which is 100% specificity at 0.47 probability and 98.04% specificity at 0.105 probability, respectively.

All models are implemented in python using the TensorFlow framework. We performed our experiments using NVIDIA GeForce GTX Titan X GPU with 12 GB of memory. The running time for classification of each patients varies from 2.26 to 22.98 seconds, depending on size of the CT image.

IV. Conclusion

In this paper, we propose a comprehensive end-to-end deep learning pipeline to quantify emphysema distribution and extent in thoracic CT images by classifying five common lung tissue patterns involved in most restrictive and obstructive differential lung disease diagnosis.

The proposed pipeline can be readily employed to characterize the presence and progression of emphysema, which makes this pipeline suitable for longitudinal analysis to evaluate time-trends for emphysema occurrence and grading.

References

[1] L. Fernandes, N. Gulati, A. M. Mesquita et al., “Quantification of Emphysema in Chronic Obstructive Pulmonary Disease by Volumetric Computed Tomography of Lung,” Indian J Chest Dis Allied Sci, vol. 57, pp. 155-60, 2015.

[2] H. Omori, K. Fujimoto, and T. Katoh, “Computed-tomography findings of emphysema: correlation with spirometric values,” Curr Opin Pulm Med, vol. 14, pp. 110-4, 2008.

[3] M. M. Wille, L. H. Thomsen, J. Petersen et al., “Visual assessment of early emphysema and interstitial abnormalities on CT is useful in lung cancer risk analysis,” Eur Radiol, vol. 26, pp. 487-94, 2016.

[4] D. A. Lynch, J. H. Austin, J. C. Hogg et al., “CT-Definable Subtypes of Chronic Obstructive Pulmonary Disease: A Statement of the Fleischner Society,” Radiology, vol. 277, pp. 192-205, 2015.

[5] L. A. Hohberger, D. R. Schroeder, B. J. Bartholmai et al., “Correlation of Regional Emphysema and Lung Cancer: A Lung Tissue Research Consortium-Based Study,” Journal of Thoracic Oncology, vol. 9, pp. 639-645, 2014.

[6] E. A. Hoffman, J. M. Reinhardt, M. Sonka et al., “Characterization of the interstitial lung diseases via density-based and texture-based analysis of computed tomography images of lung structure and function,” Acad Radiol, vol. 10, pp. 1104-18, 2003.

[7] S. P. Bhatt, G. R. Washko, E. A. Hoffman et al., “Imaging Advances in Chronic Obstructive Pulmonary Disease: Insights from COPDGene,” Am J Respir Crit Care Med, 2018.

[8] M. Anthimopoulos, S. Christodoulidis, L. Ebner et al., “Lung Pattern Classification for Interstitial Lung Diseases Using a Deep Convolutional Neural Network,” IEEE Transactions on Medical Imaging, vol. 35, pp. 1207-1216, 2016.

[9] M. Gao, U. Bagci, L. Lu et al., “Holistic classification of CT attenuation patterns for interstitial lung diseases via deep convolutional neural networks,” Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, vol. 6, pp. 1-6, 2018.

[10] S. N. Ørting, J. Petersen, L. H. Thomsen et al., "Detecting emphysema with multiple instance learning," 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), 2018, pp. 510-513.

[11] S. Christodoulidis, M. Anthimopoulos, L. Ebner et al., “Multisource Transfer Learning With Convolutional Neural Networks for Lung Pattern Analysis,” IEEE Journal of Biomedical and Health Informatics, vol. 21, pp. 76-84, 2017.

[12] D. Bermejo-Peláez, R. S. J. Estepar, and M. J. Ledesma-Carbayo, "Emphysema classification using a multi-view convolutional network," 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), 2018, pp. 519-522.

[13] M. Negahdar, D. Beymer, and T. Syeda-Mahmood, "Automated volumetric lung segmentation of thoracic CT images using fully convolutional neural network," SPIE Medical Imaging, 2018, vol. 10575.

[14] A. Depeursinge, A. Vargas, A. Platon et al., “Building a reference multimedia database for interstitial lung diseases,” Computerized Medical Imaging and Graphics, vol. 36, pp. 227-238, 2012.

Table I: Performance of the proposed pipeline to estimate emphysema extent and grade versus annotated images for 11 patients of LTRC dataset. Second column shows the number of generated patches for each patient. Third, fourth, and fifth columns show emphysema type, percentage of emphysematous volume of the lung and emphysema grade, respectively, in annotated images. Sixth, seventh, and eighth columns show the number of emphysema patches, percentage of emphysematous volume of the lung and emphysema grade, respectively, estimated by the proposed pipeline.

	Generated patches	Emphysema type	Emphysematous area%	Emphysema grade	Predicted emphysema patches	Estimated emphysematous area %	Estimated emphysema grade
Pt 1	**12591**	**No emphysema**	**0%**	**0**	**17**	**0.13%**	**0**
Pt 2	**20467**	**Centrilobular**	**57.3%**	**4**	**10462**	**51.12%**	**4**
Pt 3	**1288**	**No emphysema**	**0%**	**0**	**0**	**0%**	**0**
Pt 4	**11547**	**Paraseptal**	**42.6%**	**3**	**5271**	**45.65%**	**3**
Pt 5	**11572**	**Centrilobular**	**2.3%**	**1**	**371**	**3.2%**	**1**
Pt 6	**6410**	**Paraseptal**	**0.5%**	**1**	**123**	**1.91%**	**1**
Pt 7	**10620**	**Centrilobular**	**57.7%**	**4**	**6767**	**63.72%**	**4**
Pt 8	**11029**	**Panlobular**	**55.8%**	**4**	**7061**	**64.02%**	**4**
Pt 9	**11567**	**Centrilobular**	**40.6%**	**3**	**2520**	**21.79%**	**2**
Pt 10	**14198**	**Centrilobular**	**44.9%**	**3**	**8773**	**61.79**	**4**
Pt 11	**20134**	**Centrilobular**	**55.5%**	**4**	**11367**	**56.46**	**4**
Total	***131423***						

[15] B. Bartholmai, R. Karwoski, V. Zavaletta et al., "The Lung Tissue Research Consortium: An extensive open database containing histological, clinical, and radiological data to study chronic lung disease," 2006.

[16] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," 3D Vision (3DV), 2016 Fourth International Conference on, 2016, pp. 565-571.

[17] M. Negahdar, and D. Beymer, "Lung Tissue Characterization for Emphysema Differential Diagnosis using Deep Convolutional Neural Networks," SPIE Medical Imaging, 2019, vol. 10950.

[18] A. Krizhevsky, I. Sutskever, and G. E. Hinton, [ImageNet classification with deep convolutional neural networks] Curran Associates Inc., Lake Tahoe, Nevada(2012).

[19] C. Szegedy, L. Wei, J. Yangqing et al., "Going deeper with convolutions," 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 1-9.

[20] K. He, X. Zhang, S. Ren et al., "Deep Residual Learning for Image Recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770-778.

[21] G. Huang, Z. Liu, L. van der Maaten et al., [Densely Connected Convolutional Networks], 2261-2269 (2017).

[22] M. M. W. Wille, L. H. Thomsen, A. Dirksen et al., "Emphysema progression is visually detectable in low-dose CT in continuous but not in former smokers," European Radiology, vol. 24, pp. 2692-2699, 2014.