PHT-BOT:DEEP-LEARNING BASED SYSTEM FOR AUTOMATIC RISK STRATIFICATION OF COPD PATIENTS BASED UPON SIGNS OF PULMONARY HYPERTENSION

David Chettrit[a], Orna Bregman Amitai[a], Itamar Tamir[b], Amir Bar[a], and Eldad Elnekave[a]

[a]Zebra Medical Vision Ltd
[b]Rabin Medical Center

ABSTRACT

Chronic Obstructive Pulmonary Disease (COPD) is a leading cause of morbidity and mortality worldwide. Identifying those at highest risk of deterioration would allow more effective distribution of preventative and surveillance resources. Secondary pulmonary hypertension is a manifestation of advanced COPD, which can be reliably diagnosed by the main Pulmonary Artery (PA) to Ascending Aorta (Ao) ratio. In effect, a PA diameter to Ao diameter ratio of greater than 1 has been demonstrated to be a reliable marker of increased pulmonary arterial pressure. Although clinically valuable and readily visualized, the manual assessment of the PA and the Ao diameters is time consuming and under-reported.

The present study describes a non invasive method to measure the diameters of both the Ao and the PA from contrast-enhanced chest Computed Tomography (CT). The solution applies deep learning techniques in order to select the correct axial slice to measure, and to segment both arteries. The system achieves test Pearson correlation coefficient scores of 93% for the Ao and 92% for the PA. To the best of our knowledge, it is the first such fully automated solution.

Keywords: Deep Learning, Chest Computed Tomography, Pulmonary Hypertension, Chronic Obstructive Pulmonary Disease, Computer Aided Diagnosis.

1. INTRODUCTION

Chronic Obstructive Pulmonary Disease (COPD) affects between 4-8% of the western population. The disease course is variable, with some individuals virtually asymptomatic and others prone to repeated exacerbations and functional decline. Pulmonary hypertension (PHT) defines increased blood pressure within the pulmonary arteries. The main PA is anatomically positioned adjacent to the Ao. A PA to Ao ratio of greater than 1 has been demonstrated to be a reliable marker of increased pulmonary arterial pressure.[1,2] Enlargement of the PA or a PA to Ao ratio of greater than 1, was also shown to be among the most robust predictors of COPD exacerbation.[3–5] This is clinically intuitive, as the presence of secondary pulmonary hypertension is a sign of advanced COPD.

Pulmonary hypertension can be diagnosed via direct intra-vascular pressure measurements. It can also be inferred from radiographic findings, specifically enlargement of the pulmonary arteries, which can be rapidly assessed by comparing the diameters of the PA to Ao. This assessment of the Ao and PA diameter is time consuming and often overlooked by radiologists. Our system provides automatic measurements of the main PA and the Ao. In the COPD population, this metric may be utilized as input for population health risk-stratification. In general, these measurements may identify those who could benefit from further evaluation for pulmonary hypertension and, if the Ao is enlarged, screened for aortic aneurysms. The described solution has been developed for scans with intravenous contrast, where the blood vessels in the scan are enhanced, thus facilitating their measurements.

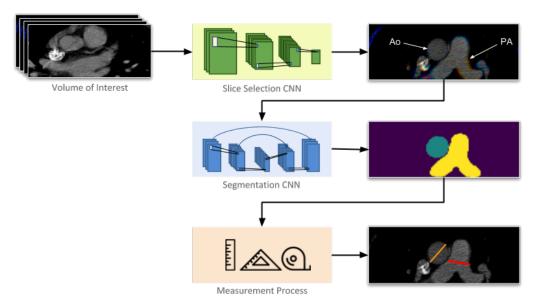Send correspondence to D.C.
E-mail: chettrit.david@zebra-med.com

Figure 1: The core system flow with its three consecutive steps of and their respective output. Given the volume of interest, the slice selection CNN is used to select the correct slice for the measurements. The two arteries are labelled in white in the chosen slice, for the reader's convenience. The chosen slice segmentation map is then generated using a second CNN. Finally, image processing techniques are used to measure the diameters, as shown in orange for the Ao and red for the PA.

The PA and Ao diameters are measured at the main PA bifurcation, on a unique axial slice. The PA is measured above its bifurcation at its maximum diameter, and the Ao is measured at the same level.[6,7] The system described here measures the arteries in this specific way in order to simulate as accurately as possible the radiologist's process.

2. PREVIOUS WORK

To the best of our knowledge, no deep learning based solution for fully automated measurement of the PA to Ao ratio on contrast chest CT scan has ever been presented. Previous work included semi-automated solutions based on traditional computer vision techniques to measure either both arteries, or just the PA diameter.[8,9] Yiting et al.[10] developed an automated solution on low-dose non contrast CT scan solution using an anatomy label map as a seed to segment the arteries of interests. The solution we propose was trained to precisely locate the correct axial slice for measurement, and to segment it automatically using a convolutional neural network (CNN). CNNs have become popular in natural image classification problems since Krizhevsky et al.[11] presented AlexNet, a convolutional neural network which outperformed previous traditional models on natural image classification problems. Since then, CNNs have been further developed and deeper architectures were introduced. Among others, UNet architectures for CNN introduced by Ronneberger et al.[12] were designed for segmentation tasks, especially on bio-medical image segmentation tasks. These models were proven to be reliable, as well as efficient tools in medical imaging, as summarized by Shen and his team.[13] Wang et al.[14] demonstrated they could be used to automatically detect pathologies on medical X-ray, and Shin et al.[15] provided a similar proof on computed tomography scans. Laserson et al.[16] also demonstrated the power of CNNs to detect 40 pathologies on chest X-rays. They also reported a higher rate of radiologist-algorithm agreement, than radiologist-radiologist agreement.

Table 1: Data properties distribution

| Data type | Train & Validation | | Test |
	Measurement	Segmentation	Measurement
size (studies number)	600	685	288
female ratio (%)	66	49	46
age mean (standard deviation)	63.1 (19.7)	66.0 (17.4)	62.0 (16.8)
cases with PA to Ao > 1 (%)	12.8	-	9.0

Distribution of various properties of interest inside the train, validation and test data sets: size, female ratio, age mean (with standard deviation) and ratio of cases with PA to Ao > 1. The statistics on the train and validation sets are presented together in the first two columns denoting the type of data set (measurement or segmentation).

3. MATERIAL

3.1 Data

Altogether, 1,285 Chest CT studies were used for training and validation (81.7% of our data set), the remaining 288 studies were used for testing (18.3%). 600 Chest CT studies were manually annotated by three experienced radiologists. The radiologists measured both arteries diameters and also reported the measurement slice. The ground truth was set to the mean of the three radiologists measurements. Additionally, 685 studies with annotated masks per artery were used for the segmentation model training and validation. For testing and reporting our results, we used an additional set unseen during training. In this set, the exact measurement locations were marked on the scans, in order to visualize the measurement process of the radiologist, and to compare it with ours. Table 1 presents the diverse properties of our data sets. In effect, this table includes statistics regarding the age, sex, size, as well as the ratio of cases with PA to Ao ratio greater than 1, within our training, validation and test sets. Our data set was labelled differently, since it is used for classification, segmentation and measurements tasks. The segmentation set, for example, does not contain ground truth measurements of both arteries, so no statistics are given regarding the distribution of cases with PA to Ao ratio > 1 within this set.

3.2 Preprocessing

Our system uses thresholding and connected components analysis, as well as morphological operations in order to detect the lungs, the trachea and the carina. Based on the location of the lungs and the carina, a volume of interest around the main PA bifurcation is defined. This volume usually comprises between 20 to 50 slices, depending on the slice thickness and increment of the series. In order to focus on the two arteries of interest only, their expected Hounsfield Unit (HU) range was considered. In effect, we applied a windowing and converted all values of the volume of interest to values between 0 and 1. The windowing range is set between -50 HU and 450 HU, thus using a windowing with the following parameters 200 HU and 250 HU for the window center (WC) and the window length (WL), respectively.

4. METHODS

The core part of the system is made of three consecutive steps. The localization of the slice containing the main PA bifurcation, the arteries segmentation and their respective measurement. The core system flow is shown in figure 1, and further described in details in this section, per component.

4.1 Slice Selection

In order to localize the slice containing the main PA bifurcation within the volume of interest, we trained a shallow CNN to recognize its properties. The CNN was given an equal number of slices suitable for the measurements and other randomly sampled slices from the volume of interest, in order to learn the differences between them. Given a slice as input, the model returns a number between 0 and 1.

The CNN is composed of a few convolutions and max pooling layers, followed by a few fully connected layers. In effect, it consists of three 3x3 convolution-batch normalization-relu-max pooling blocks, followed by a fully connected-dropout-relu-fully connected block, with a sigmoid activation at the end. This model was trained using the binary cross entropy loss function. The CNN receives as input a resized slice of the volume of interest, as well as its two closest neighbors, namely the slices below and above. It then returns the score for the middle slice. The system chooses a unique slice by selecting the slice with the highest of such scores over the entire volume of interest.

This model was trained and validated on the 600 studies containing the exact measurements of the arteries. The ground truth was set to the slice measured by a single radiologist. Statistics on the test set showed that 78.8% of the slices chosen by the network are within less than 5mm from the annotated scan and 94% within 10mm.
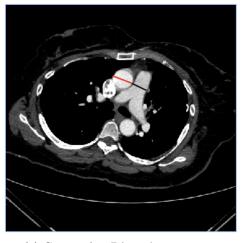
4.2 Slice Segmentation

Once a specific slice is selected, a fully convolutional CNN is used to segment both arteries. This network's architecture is a U-Net variant with skip connections, using VGG16-like[17] CNN as its encoder. We have limited it to three layers depth to avoid over-fitting of the model. In effect, this encoder is made of only two convolution-convolution-2x2 max pooling blocks followed by convolution-convolution-convolution-2x2 max pooling block, where all convolutions are 3x3, padded and contain rectified linear unit (relu) activations. Additionally, convolutions are added at the bottleneck of the encoder: a 7x7 convolution-relu-dropout and a 1x1 convolution-relu-dropout block. The up-sampling path consists of multiple transposed convolutions: a 1x1 convolution, followed by a 4x4 and a 16x16 transposed convolutions. A skip connection is used to make sure information from the down-sampling path is preserved in the up-sampling process. Indeed, the result of the first transposed convolution is combined with the result of the second convolutional block of the down-sampling path, by summing the resulting tensors. The up-sampling brings back the image to its original size. The model is trained with the sparse cross entropy loss.

The segmentation data was built in an iterative process. During training, the model was fed with manually annotated masks and semi-automatic ones (only 15% were manually tagged). In effect, we gathered additional data for training by running the initial version of the model on unknown data set, and by visually reviewing the segmentation results. As input, the network takes three different windowed versions of the slice, in order to cope with various contrast phases of the scan. These slices are centered and zero-padded to fit a fixed 512x512 input. The middle slice uses the same windowing covering the wide range of common arteries HU intensities, so a WC of 200 HU, and WL of 250 HU. The second windowing is designed for delayed venous phase, and is defined with the following parameters: WC = 100 HU, WL = 150 HU. The third windowing is designed for an earlier phase of contrast, whereby both PA and Ao are enhanced and has the following parameters WC = 375 HU, WL = 175 HU. The model returns a segmentation mask containing the three classes of interest Ao, PA and background.

4.3 Arterial Diameter Measurement

The last step in the system's flow is to measure the arteries diameter. Shape analysis and morphological operations are used in this component to measure them correctly. The segmentation from the previous step is checked, since it is expected to present a set of properties. In effect, the segmentation output is to contain exactly two distinguishable components, for each segmented artery. The Ao is to be circular, therefore its maximum eccentricity is set to 0.55. The PA bounding box to filled area ratio should be limited, since a Y-shaped PA is expected. Also, the cross section size of both arteries is expected within reasonable boundaries of (256 to 500 mm^2), and the PA centroid is expected to the right of the Ao. These various checks caused approximately 21% yield reduction, but led to higher performance and accuracy of the system.

Assuming a correct segmentation, the measurement process is started. Given that the Ao is assumed to be ellipsoidal, its measurement is relatively easy. Using the second order central moments to construct a covariance matrix of the image, one can fit an ellipse to its mask. We can then average the major and minor axis of the ellipse to provide an accurate estimate of its diameter. On the other hand, the PA can have very different shapes, and its measurement is therefore more complex. The main PA bifurcation is first localized on the skeleton of the PA binary mask using connected component analysis. Then, a stable interval of measurement above the

(a) Case with a PA to Ao ratio < 1.   (b) Case with a PA to Ao ratio > 1.

Figure 2: Sample of the system visual results. The measurements are shown in red and black for the Ao and PA, respectively.

bifurcation on the PA trunk is considered, at the same height as the Ao's center (following the measurement instructions given in section 1). By stable intervals here, we are looking into sections of the PA with minimal change in size, in order to measure properly measure the PA diameter. The interval values are finally averaged to compute the PA diameter.

5. EXPERIMENTS & RESULTS

Ground truth measurements were established by a radiologist on a test set of 288 CT studies, and compared to the algorithm outputs. Figure 3 presents multiple plots comparing the system's measurements to the ground truth reference measurements. Each compliant series within the study was automatically measured by the system. Cases with significant motion artifacts, or containing metallic artifacts were excluded from these experiments. The results show Pearson correlation coefficient scores of 93% for the Ao and 92% for the PA. Bland-Altman plots showing the limits of agreement for the measurements of the two arteries are also presented. The Bland-Altman plots also show a mean difference and standard deviation of -0.94 mm and 1.6 mm for the Ao, and of -0.86 mm and 2.03 mm for the PA. Both negative biases are caused by the system tending to slightly overestimate the arteries diameter. This can be explained by the difficulty of the segmentation model to classify pixels at the arteries borders.

As demonstrated on the confusion matrix in table 2, 91.9% of the test cases were correctly risk stratified, based on the PA to Ao ratio > 1 definition for PHT. Our solution's precision (positive predictive value) is 80.3%, meaning that a patient with a positive algorithm answer has 80.3% chance of having PHT. The system's sensitivity is 64.6%, whereas its specificity reaches 97.0%. The system's high specificity guarantees low false positive rates, thus making it ideal for screening tasks on large data sets.

Over and above our results statistics, we asked radiologists to review the measurements of the algorithm as plotted on top of the measured axial slice on 410 additional test CT scans, and to label the results as acceptable or not. In addition, we sought to quantify how frequently the algorithm measured the arteries on the correct slice. The algorithm measurements were judged acceptable and in the correct location for 99.05% and 99.76% of the cases, respectively. The unacceptable cases included those with metallic artifacts or severe motion artifacts. These results demonstrate high agreement of the radiologists with the algorithm. A sample of the system visual results on cases with a physiologic and large PA to Ao ratio are shown in figure 2.

Concerning the performance, the system average run time per study was also evaluated. Using a Intel Xeon Processor E5-2695 v4 machine, which is a similar hardware to the one on radiologists workstation, the system

Table 2: Test set confusion matrix

		Reference	
		no PHT	PHT
Algorithm	no PHT	81.7%	5.6%
	PHT	2.5%	10.2%

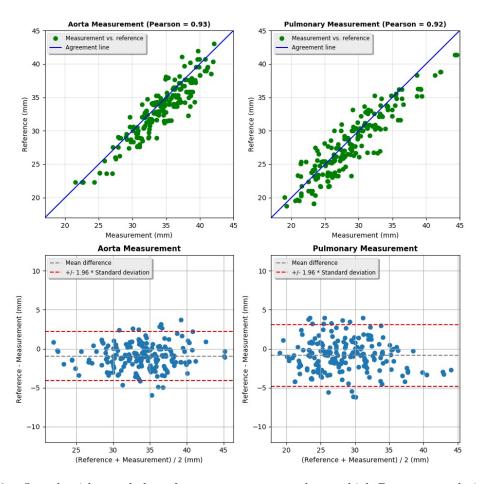Confusion matrix of the algorithm versus the reference on the test set.

Figure 3: Our algorithm and the reference measurements have a high Pearson correlation score: 93% for the Ao and 92% for the PA. The top two plots show our measurements versus the reference per artery, knowing that perfect agreement is represented by the blue diagonal line. The bottom two plots display the Bland Altman plots per artery, with the mean difference plotted via the gray dotted line. The limits of agreement are also plotted in red (-4.10 mm, 2.22 mm) for the Ao and (-4.84 mm, 3.12 mm) for the PA.

run time is 70 seconds per study on average. Using a GTX 1081Ti Graphics Processing Unit, the system run time drops to 22 seconds per study on average.

6. SUMMARY

The present solution provides automatic measurements of the Ao and PA from contrast-enhanced chest CT scans. It applies deep learning techniques as part of its flow in order to measure both arteries in the appropriate axial

location. On top of the measurement values, the system also returns visual results, thus providing interpretable feedback to its operator. Evidence of PHT may be derived from the absolute PA diameter or the PA to Ao ratio, and detection of aortic aneurysms could be a secondary gain. Detecting PHT also allows more effective risk stratification of individuals with COPD, which could improve our ability to direct preventative and health surveillance resources. Thus preventing hospital admission and the clinical sequelae of COPD exacerbation. Knowing that radiologists usually do not measure both arteries while analyzing a CT scan, the described solution provides the operator with clinically valuable information with minimal additional time cost.

To the best of our knowledge, the system is the first such fully automated solution potentially allowing screening for COPD on large CT data sets. Future work will expand current techniques to apply to non-contrast CT cases.

7. ACKNOWLEDGMENTS

REFERENCES

[1] Iyer, A. S., Wells, J. M., Vishin, S., Bhat, S. P., Wille, K. M., and Dransfield, M. T., "Ct scan-measured pulmonary artery to aorta ratio and echocardiography for detecting pulmonary hypertension in severe copd," *Chest* **145**(4), 824–832 (2013).

[2] Corson, N., Armato, G., Labby, Z. E., Straus, C., Starkey, A., and Gomberg-Maitland, M., "Ct-based pulmonary artery measurements for the assessment of pulmonary hypertension," *Acad. Radiol.* **21**(4), 523–530 (2014).

[3] Adir, Y. and Harari, S., "Pulmonary hypertension associated with chronic obstructive lung disease and idiopathic pulmonary fibrosis," *Curr. Opin. Pulm. Med.* **20**(5), 414–420 (2014).

[4] Karakus, G., Kammerlander, A. A., Aschauer, S., Marzluf, B., Zotter-Tufaro, C., Bachmann, A., Degirmencioglu, A., Duca, F., Babayev, S., Pfaffenberger, S., Bonderman, D., and Mascherbauer, J., "Pulmonary artery to aorta ratio for the detection of pulmonary hypertension: cardiovascular magnetic resonance and invasive hemodynamics in heart failure with preserved ejection fraction," *J. Cardiovasc. Magn. Reson.* **17**, 79 (2015).

[5] Wells, J. M., Washko, G., Han, M. K., Abbas, N., Nath, H., Mamary, A. J., Regan, E., Bailey, W. C., Martinez, F. J., Westfall, E., Beaty, T. H., Curran-Everett, D., Curtis, J. L., Hokanson, J. E., Lynch, D., Make, B., Crapo, J. D., Silverman, E. K., Bowler, R. P., and Dransfield, M. T., "Pulmonary arterial enlargement and acute exacerbations of copd," *N. Engl. J. Med.* **367**(10), 913–921 (2012).

[6] Ussavarungsi, K., Whitlock, J., Lundy, T., Carabenciov, I. D., Burger, C. D., and Lee, A. S., "The significance of pulmonary artery size in pulmonary hypertension," *Diseases* **2**, 243–259 (2014).

[7] Lee, S. H., Kim, Y. J., Lee, H. J., Kim, H., Kang, Y. A., Park, M. S., Kim, Y. S., Kim, S. K., Chang, J., and Jung, J. Y., "Comparison of ct-determined pulmonary artery diameter, aortic diameter, and their ratio in healthy and diverse clinical conditions," *PLOS One* **10**(5) (2015).

[8] Linguraru, M. G., Pura, J. A., Van Uitert, R. L., Mukherjee, N., Summers, R. M., Minniti, C., Gladwin, M. T., Kato, G., Machado, R. F., and Wood, B. J., "Segmentation and quantification of pulmonary artery for noninvasive ct assessment of sickle cell secondary pulmonary hypertension," *Med. Phys.* **37**(4), 1522–1532 (2010).

[9] Moses, D., Sammut, C., and Zrimec, T., "Automatic segmentation and analysis of the main pulmonary artery on standard post-contrast ct studies using iterative erosion and dilation," *Int. J. Comput. Assist. Radiol. Surg.* **11**(3), 381–395 (2016).

[10] Xie, Y., Mingzhu, L., Yankelevitz, D. F., Henschke, C., and Reeves, A. P., "Automated measurement of pulmonary artery in low-dose non-contrast chest ct images," *Proc. SPIE* **9414**, 3462–3471 (2015).

[11] Krizhevsky, A., Sutskever, I., and Hinton, G. E., "Imagenet classification with deep convolutional neural networks," in [*Advances in Neural Information Processing Systems 25*], 1097–1105 (2012).

[12] Ronneberger, O., Fischer, P., and Brox, T., "U-net: Convolutional networks for biomedical image segmentation," *MICCAI 2015. Lecture Notes in Computer Science* **9351** (2015).

[13] Shen, D., Wu, G., and Suk, H., "Deep learning in medical image analysis," *Annu. Rev. Biomed. Eng.* **19**, 221–248 (2017).

[14] Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., and Summers, R. M., "Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* , 3462–3471 (2017).

[15] Shin, H., Lu, L amd Kim, L., Seff, A., Yao, J., and Summers, R. M., "Inter-leaved textimage deep mining on a large-scale radiology database for automated image," *Journal of Machine Learning Research* **17**(107), 1–31 (2016).

[16] Laserson, J., Lantsman, C. D., Cohen-Sfady, M., Tamir, I., Goz, E., Brestel, C., Bar, S., Atar, M., and Elnekave, E., "Textray: Mining clinical reports to gain a broad understanding of chest x-rays," in [*Medical Image Computing and Computer Assisted Intervention - MICCAI 2018 - 21st International Conference, Proceedings, Part II*], 553–561 (2018).

[17] Simonyan, K. and Zissermann, A., "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556* (2014).