Function Follows Form: Regression from Complete Thoracic Computed Tomography Scans

Max Argus[1], Cornelia Schaefer-Prokop,[2] David A. Lynch[3],
Bram van Ginneken[1],

1 Diagnostic Image Analysis Group at Radboud University Medical Center,
Nijmegen, The Netherlands.
2 Meander Medical Center, Amersfoort, The Netherlands.
3 Jewish National Health, Denver, CO., USA.

Abstract

Chronic Obstructive Pulmonary Disease (COPD) is a leading cause of morbidity and mortality. While COPD diagnosis is based on lung function tests, early stages and progression of different aspects of the disease can be visible and quantitatively assessed on computed tomography (CT) scans. Many studies have been published that quantify imaging biomarkers related to COPD. In this paper we present a convolutional neural network that directly computes visual emphysema scores and predicts the outcome of lung function tests for 195 CT scans from the COPDGene study. Contrary to previous work, the proposed method does not encode any specific prior knowledge about what to quantify, but it is trained end-to-end with a set of 1424 CT scans for which the output parameters were available. The network provided state-of-the-art results for these tasks: Visual emphysema scores are comparable to those assessed by trained human observers; COPD diagnosis from estimated lung function reaches an area under the ROC curve of 0.94, outperforming prior art. The method is easily generalizable to other situations where information from whole scans needs to be summarized in single quantities.

1 Introduction

Volumetric medical scans contain a wealth of information, spatially encoded in a very large number of voxels, in the order of 10^9 for computed tomography (CT) scans. Often clinicians are interested in single quantities that represent the status of the patient. Ideally, CT scans would allow such measurements in an objective and repeatable manner. From a machine learning perspective, reducing 10^9 measurements to a single number is a daunting task. As a consequence,

automated systems that have been developed to carry out such tasks typically employ a sequence of steps: data is first preprocessed, structures that are considered most relevant are segmented, within these structures certain regions are identified where specific measurements are performed, and these measurements are aggregated into the final quantity to be extracted. Clearly, such systems are highly task specific and result from a large number of dedicated design decisions.

In this work we pursue an orthogonal approach and present a system, based on a convolutional network, that is intended to generically carry out the task of translating a volumetric scan directly into a single measurement. The system is trained end-to-end using a set of pairs of scans (the input) and the quantity to be derived from the scans (the output). No specific medical knowledge about what in the scan is important to derive the desired output is encoded into the system. We therefore refer to our system with the general name SCAN2NUM.

We apply SCAN2NUM to thoracic computed tomography (CT) scans. We present results for various tasks related to chronic obstructive pulmonary disease (COPD). This is currently the fourth leading cause of death worldwide, and expected to be the third leading cause of death in 2020. In COPD patients chronic inflammation of the airways and the lungs, usually due to smoking, leads to structural changes in the airways and lung parenchyma. The airflow limitation in COPD is assessed spirometrically by a combination of forced expiratory volume in 1 second (FEV_1) and the forced vital capacity (FVC). From these quantities the GOLD score is computed, with which COPD is diagnosed and the severity of the disease is quantified.

Where spirometry measures global functional aspects of the disease, CT spatially resolves the pathophysiological processes and structural damage. In the CT scans airway wall thickening can be observed, as well as emphysema resulting from degradation of lung tissue through the enlargement and destruction of alveoli. CT shows both the severity and extent of different emphysematous patterns, such as centrilobular, parenchymal and paraseptal emphysema. This information is reported by radiologists to inform treatment decisions [1]. Visual assessment has, however, been found to have a large inter-reader variability when reporting these quantities [2].

Ideally, analysis of CT scans would provide sensitive and repeatable measurements of disease progression. From CT scans it may be possible to quantify not only physical progression, but also the functional progression by training models to predict functional quantities. Such appearance based measurements could be further used for screening for COPD, for monitoring and for measuring response to treatment. Imaging biomarkers that correlate with progression are increasingly attracting attention, as a way to measure the effects of different treatment options [3, 4].

To demonstrate the potential of SCAN2NUM, we apply the network in this study to various quantitative tasks related to COPD. We trained it to predict standardized parenchymal emphysema scores, and to assess the lung function test scores FEV_1/FVC and $FEV_1\%$predicted directly from the CT scan. These two scores can be used to diagnose COPD and determine the GOLD stage of the patient.

We believe that this network can also be applied to other, similar, problems and hopefully is a step in the direction of more end-to-end optimization for CT.

1.1 Related Work

In this section we briefly review prior work on the tasks we address in this study, namely emphysema quantification and deriving estimates of lung function measurements from CT scans, and we report on related work on deriving measurements from 3D scans using deep networks.

Emphysema quantification can be performed using a densitometric approach or a texture based approach. In the standard densitometric analysis of emphysema an emphysema score is computed as the percentage of lung tissue, excluding airways, with a density less than a given threshold value [5]. This method has been applied to the COPDGene dataset in [6].

As this standard method suffers from lack of reproducibility a number of correction methodologies were introduced, such as volume correction [7] and kernel hardness correction [8].

As an alternative to purely densitometric approaches texture based approaches have also been introduced. The prior work in [9], [10], [11], [12], [13], [14] usually starts by extracting local, hand crafted features from Regions of Interest(RoI). These features are then used to train a classifier to classify the RoIs into the previously annotated lung tissue types. Sometimes local features are also learned based on manual RoI annotations.

Clustering avoids manual RoI annotation and is applied to emphysema subtyping in [15], as does multiple instance learning, which is applied to the problem of COPD classification in [16]. This approach, nevertheless, requires hand crafted features as a basis.

Prior work in predicting lung function from CT scans also relies on hand crafted features. These features are often quite coarse, relying on organ level thresholds, medians and volumes as can be seen in [17].

Transitioning from hand crafted features to learned features is an important step as the latter can often be more expressive and efficacious. Similarly not relying on RoIs is also important as, in many cases, the functional quantities of interest cannot be directly associated with visual patterns.

Scan-level CNN models have also recently been used for longevity prediction from CT scans in [18]. This approach makes use of a 3D CNN model, but therefore requires aggressive down-sampling of the scans. We instead advocate a late fusion architecture, as this allows processing the scans at full resolution which better enables texture based emphysema classification.

2 Materials and methods

2.1 Subjects and Image Acquisition

A major hurdle to applying large models, such as convolutional networks, to make scan level predictions is that a large number of scan-level examples are required to avoid overfitting. The chest CT scans used in this study were obtained from the multi-center COPDGene Study. This study included 10,192 current or former smokers and 108 non-smokers. Details regarding the inclusion criteria and CT acquisition can be found in the COPDGene study design paper [19]. Briefly, subjects were scanned at full inspiration with 200 mAs. Thin-slice reconstructions were used. Scans come from 21 centers and a wide variety of

scanners were used. When multiple reconstruction kernels were available, we choose the softer kernel.

We used baseline CT scans from those subjects for which both the visual emphysema scores and lung function measurements were available. This yielded a set of 1,788 scans, which were randomly split into a training set of 1,424 scans a validation set of 170 scans and a test set of 195 scans, stratified according to COPD Gold stages and outcome parameters.

2.2 Assessment of Visual Emphysema Score

The large amount of inter-reader variability in emphysema classification and quantification has lead to efforts to precisely delineate different visual phenotypes [1] and create reader training and scoring procedures in order to produce consistent measures.

This was done for parenchymal emphysema, the most common type of emphysema for COPD patients. In parenchymal emphysema small lesions form at the center of the secondary lobule and in case of progression they coalesce into large areas of lung destruction. The visual scoring of emphysema follows the methodology of [20]. The severity of parenchymal emphysema was assessed by trained medical analysts according to a six point scale of: no emphysema, trace, mild, moderate, confluent or advanced destructive emphysema. The analysts scored the highest grade of emphysema that is visible on the scans, even if it was present only in one location. Weighted kappa scores for observer agreement for this scoring system ranged from 0.70 to 0.80, indicating good to excellent agreement. This quantity is referred to as the Visual Emphysema (VE) score.

3 Method: SCAN2NUM Network

The diagnostic process aims to find a single quantity that describes the state of the patient. Computationally this corresponds to the conversion of a spatially resolved image into a per-scan value. This computation is often factored into a local computation plus a global aggregation step. Many tasks such as lesion segmentation and nodule detection are focal, in the sense that decisions can be made based only on a local volume, whereas other tasks are more diffuse in nature.

While the spatial factoring of an algorithm is dictated by intrinsic attributes of the problem in focal tasks, choosing a spatial factoring for diffuse tasks is more difficult. This is especially true when applying CNNs which need to reconcile the desire for gradual spatial resolution reduction with algorithm memory usage as for training as all intermediate features of the computation need to be held in memory in order to be able to perform back-propagation.

An extreme example of spatial factoring can be seen in emphysema score computation, which also maps whole scans to single quantities, it makes local single-voxel decision by thresholding and then aggregates these by averaging.

In this work we formulate a CNN network, named SCAN2NUM, that addresses this problem for three important quantities in COPD diagnosis that describe lung function and aspect.

The architecture that we chose is based on the computer vision task of video classification [21], in which a number of architectures have been developed to
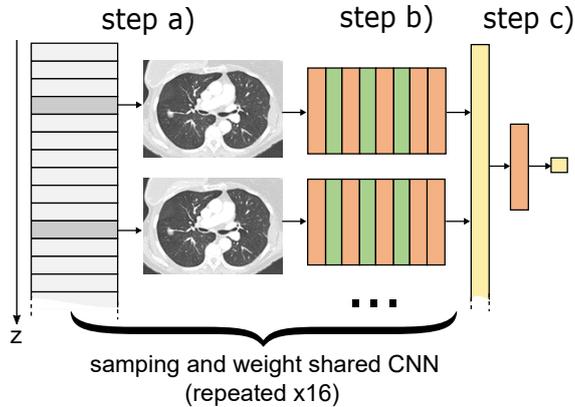
Figure 1: Late fusion SCAN2NUMnetwork for whole scan classification. The network follows three steps a) axial slice sampling, b) per-slice feature computation c) feature aggregation by summation and scoring.

solve this similar problem. For our task we chose a late-fusion architecture as these provide good classification performance. In these architectures, the same 2D network is applied to a regularly spaced subset of slices and the resulting feature vectors are averaged to aggregate the 2D slice information over the whole volume.

With this choice of model the process can be described as consisting of three main steps of preprocessing, 2D per-slice feature computation and finally feature aggregation and scoring.

In pre-processing the CT scans are cropped to the bounding box of the lung segmentation. The lung segmentations were provided in the COPDGene database and were computed using a previously published approach [22]. This step removes some of the irrelevant areas from the CT scans to help the remaining steps focus on the lung. Subsequently, regularly spaced slices are sampled from the scan.

In the per-slice computation the aim was to extract an informative feature vector, without letting the model become too large. This was done by a very narrow network in order to save memory space. As identification of emphysema depends on the local texture of the lung, slices where given to the network at their full resolution.

In the final step the per-slice features are aggregated by mean pooling of the per-slice feature vectors. This allows varying the number of slices that are considered for the score. 16 slices per scan were chosen, as this was shown to produce good results see Table 6 for details.

In the scoring step the global feature vector, representing the entire volume, is converted into a per-scan score. A L2 regression loss was chosen as need to predict continuous quantities, and our discrete VE scores were ranked. This method of training has the benefit of transmuting discrete labels into continuous ones, which may be more useful for further statistical analysis. Network architecture and training parameters can be found in Table 1.

The network chosen is inspired by the second place entry in the Kaggle Diabetic Retinopathy Detection Challenge and is described in Table 1.

Table 1: Network architecture. The network processes 16 slices which are aggregated into a single scan label.

step	layer name	channels	filter	stride	output size
a)	input	1			512
b)	conv1	32	5x5	2	254
	pool1	32	2x2	2	127
	conv2	128	5x5	2	62
	pool2	128	2x2	2	31
	conv3	256	3x3	2	15
	pool3	256	2x2	2	8
	conv4	512	3x3	2	3
	conv5	1024	3x3	1	1
c)	sum	1024			1
	fc6	1024			1
	L2 loss	1			1

We trained our models using stochastic gradient descent with a batch size of 16 examples, momentum of 0.9. Linear learning rate decay was used starting with a learning rate of 0.005 over 100k iterations. The performance on the validation set was evaluated every 500 iterations and the network with the lowest loss on the validation set was chosen as the final network. Weights are initialized using the Xavier filler [23]. Weight decay of 0.0005 was used together with dropout of 0.5 that was applied only to the fc6 layer. Data augmentation of random mirroring and per slice rotation by ±45°, sampled uniformly, was applied. During training regularly spaced slices were sampled, by varying the inital slice offset according to a uniform distribution additional data variability was achieved. All models are trained and tested with Caffe [24] on a single NVIDIA GeForce GTX 1080.

4 Results

The described network was applied to compute predictions for three quantities associated with COPD, the VE score, FEV_1/FVC and FEV_1. All results are reported for the same test set of 195 scans. The comparison of performance relative to other methods is hindered by the lack of a standardized test set. [1]

4.1 VE Classification

We started with VE score prediction. As this is a quantity that is visually assessed from the scan itself, it was the most likely to be predictable. The results are shown in Figure 2b next to a plot of the percent emphysema below the -950HU threshold as a benchmark shown in Figure 2a. In the predictions we see a good separation of classes, except between VE scores of 0–1, which correspond to absent and trace amounts of emphysema. This may be due to the fact that small amount of emphysema are more easily missed by both the slice sampling step and the feature computation step. Visual assessment of outliers is shown in

---

[1]Our test set, with individual reader scores, is available here: goo.gl/vP294U.

Figure 4. Further visual evaluation is possible by looking at the individual VE response per-slice. This should produce meaningful information as all feature aggregation operations are linear. The results of this can be seen in Figure 5. Here we see that the per slice VE score correlates with the relative presence of emphysema.

A slight systematic offset can also be seen in the results when comparing the linear and quadratic fits of the data. These deviation between these two shows that the dataset is biased to prediction the dataset mean of ~ 1.75. For further analysis of accuracy the continuous emphysema scores were discretized. This was done, without further calibration, by rounding to the nearest whole number. This results in the confusion matrix Table 2.

Table 2: Confusion Matrix of VE scores comparing visually assessed and computed scores.
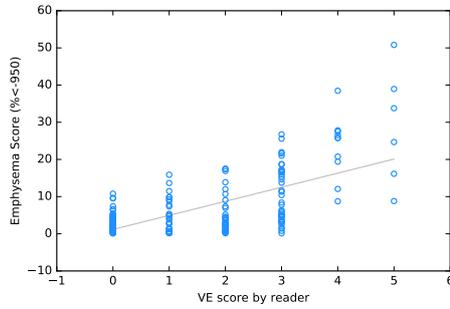
VE	Predicted					
Score	0	1	2	3	4	5
0	44	31	1			
1	10	18	2	0		
2	3	16	11	2	1	
3		1	17	15	6	1
4		1		3	5	1
5				1	4	1

The performance of the system can also be assessed with respect to the inter-reader variability. The VE scores were computed from the consensus reading of two research analysts, adjudicated by a third radiologist reading in the case of disagreements. This was assessed by comparing the inter-analyst Spearman's rank correlation of $\rho_{aa} = 0.84$, 95% CI [0.80 - 0.88] to the first-analyst system correlation of $\rho_{as} = 0.79$, 95% CI [0.74 - 0.84]. This gives $P(\rho_{aa} > \rho_{as}) = .93$. The confidence intervals and probability are computed via bootstrapping with 10k samples.

VE scores were assessed based on the location in the scan with the highest visible grade of emphysema. Even if the emphysema is very localized, this can lead to a degree of inter-reader variability in the ground truth annotation. Given these challenging conditions, the high level of performance at the task is impressive. Furthermore the clinical relevance of the results should be emphasized. The VE score has independently associated with mortality [25], indicating that this method can provide a similarly effective mortality signal. Additionally VE scores were associated with genetic markers linked to COPD susceptibility [20].

4.2 COPD Classification

The next task was FEV_1/FVC prediction. In contrast to VE scoring, this quantity is defined via a lung function test. It is correlated to the lung appearance as can be seen in the correlation with the VE score as shown in Figure 2c. The results from the automatic system can be seen in Figure 2d. From the slightly lower correlation with visual VE scores it can be seen that it is a more difficult for the system to predict this quantity visually.
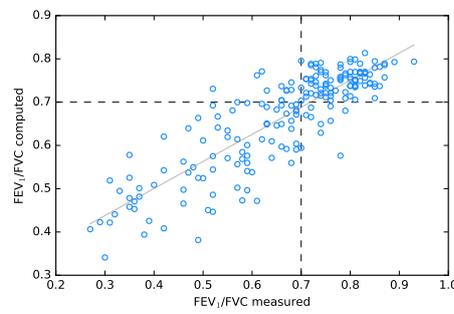
(a) %-emphysema correlation with VE score $\rho = 0.56$, 95% CI [0.46 - 0.64]
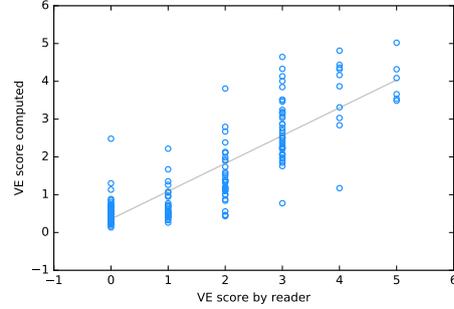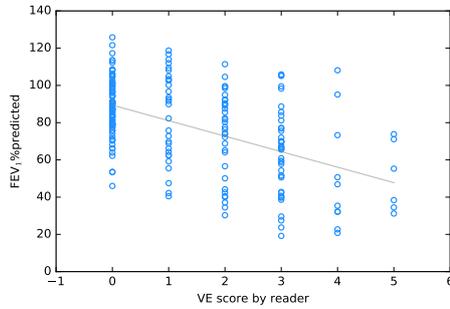
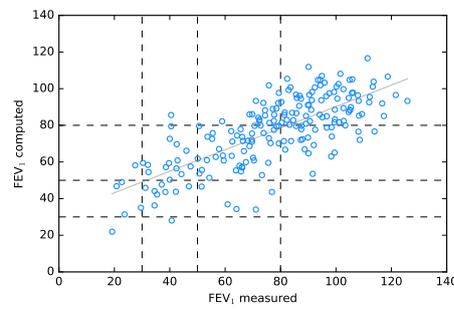(b) VE score prediction. $\rho = 0.84$, 95% CI [0.80 - 0.88]
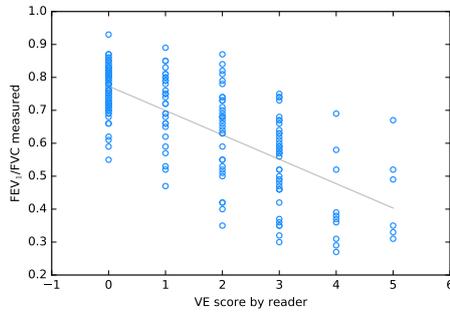
(c) FEV_1/FVC correlation with VE score. $|\rho| = 0.70$, 95% CI [0.63 - 0.75]

(d) FEV_1/FVC prediction $\rho = 0.82$, 95% CI [0.77-0.86] showing COPD threshold.

(e) FEV_1%predicted correlation with VE score. $|\rho| = 0.47$, 95% CI [0.37 - 0.56]

(f) FEV_1%predicted prediction $\rho = 0.73$, 95% CI [0.66 - 0.78] showing GOLD stage thresholds.

Figure 2: Scatter plots showing the Spearman's rank correlation of computed predictions for various quantities (right column) compared to the correlation of these quantities with chosen reference values (left column). The gray lines are linear fits, all correlations shown are for the same test set of 195 patients and have $p < .001$.

FEV$_1$/FVC is important because a patient is diagnosed with COPD if this value is below 0.7. This discretization allows a confusion matrix, Table 3, and accuracy metrics to be computed. Additionally a ROC curve can be computed by using the numerical value of FEV$_1$/FVC as a score as shown in Figure 3.

This method outperforms other CT based COPD prediction methods shown in Table 4, even without making use of additional patient information. For example age, body mass index, smoking status and pack-years of smoking history are used in [26] together with the CT derived measures of air trapping, for which an additional expiration scan is needed.

COPD	Predicted	
Present	No	Yes
No	91	10
Yes	11	83

Table 3: Confusion Matrix of COPD diagnosis comparing spirometric and computed values.

Method	AUC	95%CI
Mets [27]	0.83	0.81-0.86
Mets [26] Asymptomatic	0.83	0.80-0.87
Mets [26] Symptomatic	0.91	0.88-0.93
Ours	0.94	0.91-0.96

Table 4: Performance comparison of COPD prediction methods. (Not evaluated on the same dataset.)
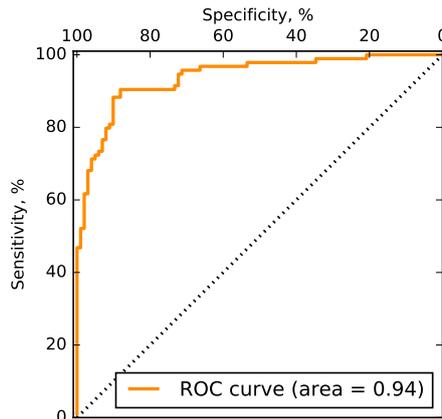

Figure 3: ROC curve for COPD classification using computed FEV$_1$/FVC values.

4.3 GOLD Stage Prediction

The next quantity we looked at was computing FEV$_1$%predicted. This is a normalized form of FEV$_1$ to account for variation in height and age. Depending on the FEV$_1$%predicted value patients have, they are stratified into different GOLD stages, which measure the severity of COPD. This quantity is even less strongly correlated with VE score, see Figure 2e, and correspondingly is harder to predict, see Figure 2f. Again we follow the same evaluation procedure of discretization and accuracy evaluation.

GOLD Stage		Predicted			
	0	1	2	3	4
Actual 0	91	3	7		
Actual 1	4	6	11		
Actual 2	6	5	28	5	
Actual 3	1		13	8	1
Actual 4			1	4	1

Table 5: Confusion Matrix of Gold Stages.

num. slices	ρ	95% CI	
8	0.78	0.72	0.82
12	0.79	0.74	0.84
16	0.82	0.77	0.86
24	0.81	0.76	0.85

Table 6: Model Variation trained on FEV_1/FVC.

4.4 Model Search

One important hyper-parameter of the model is the number of slices that are samples from the scan. In order to show how performance depends on this parameter, we tested how well FEV_1/FVC could be computed with a varying number of slices. The result of this can be seen in Table 6.

5 Discussion

We have shown that a multi-slice analysis of full 3D CT scans, trained end-to-end, can make meaningful predictions at the patient level. Exactly the same approach can be used to predict multiple outputs, without the need to encode any specific knowledge into the system about which characteristics on the scan are known to be correlated to the output parameter that we want to predict. This is in contrast to previously published methods to predict the severity of parenchymal emphysema, lung function status and COPD status. These works were based on the detection and quantification of specific local patterns in the scan, derived from lung parenchyma and the airways. One would expect that such dedicated methods would outperform the generic approach proposed here. The proposed approach could function as a baseline to benchmark specific approaches. But in fact our experimental results suggest that our system is at least equivalent or even superior to dedicated approaches. Further evaluation on common datasets is needed to analyze the full potential of deep convolutional methods versus dedicated classical approaches.

There are various strategies that could be explored to maximize performance. The most obvious suggestion is to train the systems on a larger set of CT scans. Especially for predicting lung function status, a much larger number of CTs with corresponding lung function tests are already contained in the COPDGene database. In this work we limited ourselves to the scans we could obtain from the COPDGene study for which all three output parameters we wanted to predict were available. This allowed us to perform all experiments with the same training set.

Another option is to add additional types of input data. Previous work focused on predicting COPD status included additional parameters about the patient, such as age, smoking status and history, and additional exposures are known to correlate with COPD. Such information can be inserted at various locations into the network to ensure that the entire system can still be trained
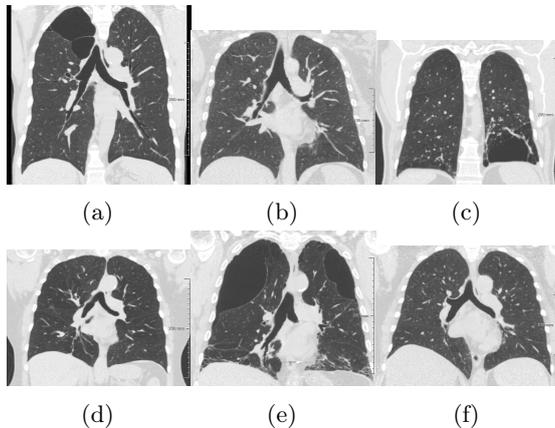
Figure 4: VE score prediction outliers. The presence of upper lobe paraseptal emphysema without parenchymal emphysema in (a)+(e) and diffuse emphysema in (d), may have lead to over-prediction, whereas lower lobe paraseptal emphysema in (c) and possible incomplete inspiration in (b)+(f) may have lead to under-prediction.

end-to-end. Additionally, expiration CT scans could be added as an additional input, and the system would analyze slices from multiple scans. This approach can also be used to extend the analysis to longitudinal scan data. If data at multiple time points would be available, prediction of clinically relevant trends, e.g. the rate of COPD progression, can be expected to be more accurate.

Slice sampling could be improved by evaluating several samples per scan and averaging the results. In the presented results, the system inspects only 16 slices from a 3D scan. One could offset the location of the start slice and run the same analysis multiple times. Furthermore, additional networks could be trained for coronal and sagittal slices and a combination of all three could be used to provide the final score. This would go some way to address a major drawback of SCAN2NUM, namely that it uses a 2D approach to analyze 3D data, and only analyzes a limited number of slices. Note that we have trained a system with 24 slices and achieved no improvement. We have also varied other aspects of the architecture. In particular, we investigated the use a Soft-Max loss for VE scoring, but this did not yield good results. Using L2 regression allowed the network to perform better. One could attempt to improve performance further in the case of tasks that produce discretized quantities by additional calibration that may compensate for prediction bias.

A potential drawback of the overall approach, that may limit wider applicability to other tasks, is the requirement of a large set of labeled scans. For this work the COPDGene trial provided these, making it an excellent setting to explore the performance of such architectures. We hope that these datasets can be used as benchmarks for evaluating architecture improvements such as those mentioned above. Better performance would allow the use of fewer scans, enabling the application of this approach to a larger set of problems.

In this work we have predicted lung function tests results from CT scans. Obviously performing lung function tests directly is a viable alternative. We like to stress that in clinical practice an enormous amount of chest CT scans
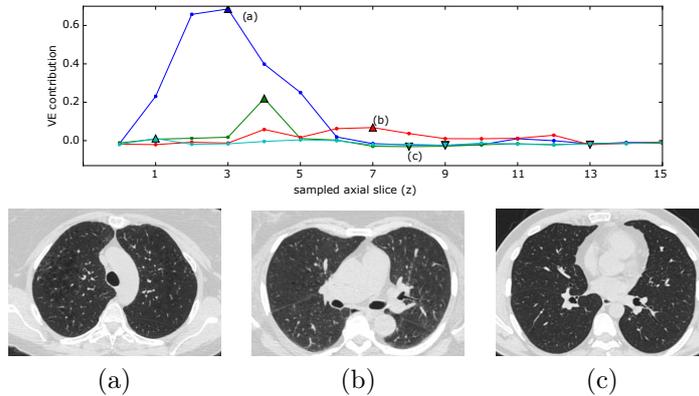
Figure 5: The plot above shows VE score contributions for each sampled axial slices for four random scans. The maximal and minimal VE response are indicated by triangles, those marked with letters are shown below. Slice a shows extensive emphysema in the center of the right lung, scan b shows a small emphysema lesion above the right fissure and slice c shows healthy slices are scored correctly. The images are best viewed zoomed in on a large computer monitor.

are acquired in subjects who are at risk for chronic lung diseases but do not undergo lung functions testing. This is especially true now CT lung cancer screening programs are being implemented. CT based surrogate lung function measurements may therefore be a good and extremely cheap way to also screen for COPD, as suggested by Mets et al. [27]. If visual surrogates become better, they could be used as a second measurement for lung function in order to decrease the variance of this test. It may be possible to use similar methods to identify which lung cancer screening participants are particularly at risk of the disease and to adjust screening frequency or follow up accordingly. A recent study indicated that risk for cardiovascular disease can be assessed by visually by radiologists [28], possibly SCAN2NUMcould achieve similar results for this task.

6 Conclusion

This paper shows that learning VE, FEV_1/FVC and FEV_1 from whole scans is possible, even with relatively simple networks. Whole scan end-to-end learning is a general method which can be applied to a large number of diagnostic and image quantification processes in radiology. We hope to have shown that the proposed method is sufficiently attractive for it to be considered for other tasks.

Acknowledgments

References

[1] Lynch DA, Austin JHM, Hogg JC, Grenier PA, Kauczor HU, Bankier AA, et al. CT-Definable Subtypes of Chronic Obstructive Pulmonary Disease: A Statement of the Fleischner Society. Radiology. 2015;277(1):192–205. doi:10.1148/radiol.2015141579.

[2] Lynch DA, Murphy JR, Crapo JD, Criner GJ, Galperin-Aizenberg M, Jacobson FL, et al. A Combined Pulmonary -Radiology Workshop for Visual Evaluation of COPD: Study Design, Chest CT Findings and Concordance with Quantitative Evaluation. COPD. 2012;9.2:151 – 159. doi:10.3109/15412555.2012.654923.

[3] Mosley J, Smith L, Dutton B. Tiotropium Bromide/Olodaterol (Stiolto Respimat): Once-Daily Combination Therapy for the Maintenance of COPD. Pharmacy and Therapeutics. 2016;41(2):97–102.

[4] Maleki-Yazdi MR, Kaelin T, Richard N, Zvarich M, Church A. Efficacy and safety of umeclidinium/vilanterol 62.5/25 mcg and tiotropium 18 mcg in chronic obstructive pulmonary disease: Results of a 24-week, randomized, controlled trial. Respiratory Medicine. 2014;108(12):1752 – 1760. doi:http://dx.doi.org/10.1016/j.rmed.2014.10.002.

[5] van Rikxoort EM, de Hoop B, van de Vorst S, Prokop M, van Ginneken B. Automatic segmentation of pulmonary segments from volumetric chest CT scans. IEEE Transactions on Medical Imaging. 2009;28:621–630. doi:10.1109/TMI.2008.2008968.

[6] Han MK, Kazerooni EA, Lynch DA, Liu LX, Murray S, Curtis JL, et al. Chronic Obstructive Pulmonary Disease Exacerbations in the COPDGene Study: Associated Radiologic Phenotypes. Radiology. 2011;261(1):274–282. doi:10.1148/radiol.11110173.

[7] Stoel BC, Putter H, Bakker ME, Dirksen A, Stockley RA, Piitulainen E, et al. Volume Correction in Computed Tomography Densitometry for Follow-up Studies on Pulmonary Emphysema. Proceedings of the American Thoracic Society. 2008;5(9):919–924. doi:10.1513/pats.200804-040QC.

[8] Gallardo-Estrella L, Lynch DA, Prokop M, Stinson D, Zach J, Judy PF, et al. Normalizing computed tomography data reconstructed with different filter kernels: effect on emphysema quantification. European Radiology. 2016;26:478–486. doi:10.1007/s00330-015-3824-y.

[9] Xu Y, Sonka M, McLennan G, Guo J, Hoffman EA. MDCT-based 3-D texture classification of emphysema and early smoking related lung pathologies. IEEE Transactions on Medical Imaging. 2006;25(4):464–475. doi:10.1109/TMI.2006.870889.

[10] Park YS, Seo JB, Kim N, Chae EJ, Oh YM, Lee SD, et al. Texture-Based Quantification of Pulmonary Emphysema on High-Resolution Computed Tomography: Comparison With Density-Based Quantification and Correlation With Pulmonary Function Test. Investigative Radiology. 2008;43(6):395–402. doi:10.1097/RLI.0b013e31816901c7.

[11] Sørensen L, Shaker SB, de Bruijne M. Quantitative Analysis of Pulmonary Emphysema Using Local Binary Patterns. IEEE Transactions on Medical Imaging. 2010;29(2):559–569. doi:10.1109/TMI.2009.2038575.

[12] Häme Y, Angelini ED, Hoffman EA, Barr RG, Laine AF. Adaptive Quantification and Longitudinal Analysis of Pulmonary Emphysema With a Hidden Markov Measure Field Model. IEEE Transactions on Medical Imaging. 2014;33(7):1527–1540. doi:10.1109/TMI.2014.2317520.

[13] Yang J, Feng X, Angelini ED, Laine AF. Texton and sparse representation based texture classification of lung parenchyma in CT images. In: 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); 2016. p. 1276–1279.

[14] Ginsburg SB, Zhao J, Humphries S, Jou S, Yagihashi K, Lynch DA, et al. Texture-based Quantification of Centrilobular Emphysema and Centrilobular Nodularity in Longitudinal CT Scans of Current and Former Smokers. Academic Radiology. 2016;23(11):1349 – 1358. doi:http://dx.doi.org/10.1016/j.acra.2016.06.002.

[15] Binder P, Batmanghelich NK, Estepar RSJ, Golland P. Unsupervised Discovery of Emphysema Subtypes in a Large Clinical Cohort. In: Wang L, Adeli E, Wang Q, Shi Y, Suk HI, editors. Machine Learning in Medical Imaging: 7th International Workshop, MLMI 2016, Held in Conjunction with MICCAI 2016, Athens, Greece, October 17, 2016, Proceedings. Cham: Springer International Publishing; 2016.

[16] Cheplygina V, Sørensen L, Tax DMJ, Pedersen JH, Loog M, d Bruijne M. Classification of COPD with Multiple Instance Learning. In: 2014 22nd International Conference on Pattern Recognition; 2014.

[17] Mets OM, Murphy K, Zanen P, Gietema HA, Lammers JW, van Ginneken B, et al. The relationship between lung function impairment and quantitative computed tomography in chronic obstructive pulmonary disease. European Radiology. 2012;22:120–128. doi:10.1007/s00330-011-2237-9.

[18] Oakden-Rayner L, Carneiro G, Bessen T, Nascimento JC, Bradley AP, Palmer LJ. Precision Radiology: Predicting longevity using feature engineering and deep learning methods in a radiomics framework. Scientific Reports. 2017;7(1648).

[19] Regan EA, Hokanson JE, Murphy JR, Make B, Lynch DA, Beaty TH, et al. Genetic Epidemiology of COPD (COPDGene) Study Design. COPD. 2010; p. 32–43. doi:10.3109/15412550903499522.

[20] Halper-Stromberg E, Cho MH, Wilson C, Nevrekar D, Crapo JD, Washko G, et al. Visual Assessment of Chest Computed Tomographic Images Is Independently Useful for Genetic Association Analysis in Studies of Chronic Obstructive Pulmonary Disease. Annals of the American Thoracic Society. 2017;14(1):33–40. doi:10.1513/AnnalsATS.201606-427OC.

[21] Karpathy A, Toderici G, Shetty S, Leung T, Sukthankar R, Fei-Fei L. Large-scale Video Classification with Convolutional Neural Networks. In: CVPR; 2014.

[22] Sluimer IC, Prokop M, van Ginneken B. Towards automated segmentation of the pathological lung in CT. IEEE Transactions on Medical Imaging. 2005;24:1025–1038. doi:10.1109/TMI.2005.851757.

[23] Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks. In: In Proceedings of the International Conference on Artificial Intelligence and Statistics (AISTATS'10). Society for Artificial Intelligence and Statistics; 2010.

[24] Jia Y, Shelhamer E, Donahue J, Karayev S, Long J, Girshick R, et al. Caffe: Convolutional Architecture for Fast Feature Embedding. arXiv preprint arXiv:14085093. 2014;.

[25] Lynch DA, Moore C, Wilson CG, Nevrekar DV, Jennermann TB, Humphries S. Visual Emphysema Pattern Using the Fleischner Society Classification System Is Independently Associated with Mortality in Cigarette Smokers. In: Annual Meeting of the Radiological Society of North America; 2017.

[26] Mets OM, Schmidt M, Buckens CF, Gondrie MJ, Isgum I, Oudkerk M, et al. Diagnosis of chronic obstructive pulmonary disease in lung cancer screening Computed Tomography scans: independent contribution of emphysema, air trapping and bronchial wall thickening. Respiratory Research. 2013;14:59. doi:10.1186/1465-9921-14-59.

[27] Mets OM, Buckens CFM, Zanen P, Isgum I, van Ginneken B, Prokop M, et al. Identification of Chronic Obstructive Pulmonary Disease in Lung Cancer Screening Computed Tomographic Scans. Journal of the American Medical Association. 2011;306:1775–1781. doi:10.1001/jama.2011.1531.

[28] Chiles C, Duan F, Gladish GW, Ravenel JG, Baginski SG, Snyder BS, et al. Association of Coronary Artery Calcification and Mortality in the National Lung Screening Trial: A Comparison of Three Scoring Methods. Radiology. 2015;276:82–90. doi:10.1148/radiol.15142062.