

©2012 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

DOI: 10.1109/TMI.2011.2164931

# Texture-Based Analysis of COPD: a Data-Driven Approach

Lauge Sørensen\*, Mads Nielsen, Pechin Lo, Haseem Ashraf, Jesper H. Pedersen, and Marleen de Bruijne

**Abstract**—This study presents a fully automatic, data-driven approach for texture-based quantitative analysis of chronic obstructive pulmonary disease (COPD) in pulmonary computed tomography (CT) images. The approach uses supervised learning where the class labels are, in contrast to previous work, based on measured lung function instead of on manually annotated regions of interest (ROIs). A quantitative measure of COPD is obtained by fusing COPD probabilities computed in ROIs within the lung fields where the individual ROI probabilities are computed using a  $k$  nearest neighbor ( $k$ NN) classifier. The distance between two ROIs in the  $k$ NN classifier is computed as the textural dissimilarity between the ROIs, where the ROI texture is described by histograms of filter responses from a multi-scale, rotation invariant Gaussian filter bank. The method was trained on 400 images from a lung cancer screening trial and subsequently applied to classify 200 independent images from the same screening trial. The texture-based measure was significantly better at discriminating between subjects with and without COPD than were the two most common quantitative measures of COPD in the literature, which are based on density. The proposed measure achieved an area under the receiver operating characteristic curve (AUC) of 0.713 whereas the best performing density measure achieved an AUC of 0.598. Further, the proposed measure is as reproducible as the density measures, and there were indications that it correlates better with lung function and is less influenced by inspiration level.

**Index Terms**—COPD, computed tomography (CT), lung, texture analysis, classification

## I. INTRODUCTION

CURRENT quantitative measures of chronic obstructive pulmonary disease (COPD) are limited in several ways. The gold standard for diagnosis of COPD is a pulmonary function test (PFT) [1]. These non-invasive measurements are

Copyright (c) 2010 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

This work was supported in part by the Danish Council for Strategic Research under the Programme Commission for Nanoscience and Technology, Biotechnology and IT, in part by the Netherlands Organization for Scientific Research, and in part by AstraZeneca, Lund, Sweden. *Asterisk indicates corresponding author.*

\*Lauge Sørensen is with the Image Group, Department of Computer Science, University of Copenhagen, Denmark. (e-mail: lauges@diku.dk).

Mads Nielsen is with the Image Group, Department of Computer Science, University of Copenhagen, Denmark and with Nordic Bioscience Imaging.

Pechin Lo is with the Image Group, Department of Computer Science, University of Copenhagen, Denmark.

Haseem Ashraf is with the Department of Radiology, Akershus University Hospital, Norway.

Jesper H. Pedersen is with Department of Thoracic Surgery, Rigshospitalet, University of Copenhagen, Denmark.

Marleen de Bruijne is with the Image Group, Department of Computer Science, University of Copenhagen, Denmark and with the Biomedical Imaging Group Rotterdam, Departments of Radiology & Medical Informatics, Erasmus MC, Rotterdam, The Netherlands.

cheap and fast to acquire but are limited by insensitivity to early stages of COPD [2] and lack of reproducibility [3]. Visual and computerized assessment in computed tomography (CT) imaging has emerged as an alternative that directly can measure the two components of COPD, namely, chronic bronchitis and emphysema. However, it is difficult to visually assess disease severity and progression. Moreover, visual assessment is subjective, time-consuming, and suffers from intra-observer and inter-observer variability [4], [5]. The most widely used computerized measures, also referred to as densitometry or quantitative CT, are the relative area of CT attenuation values below a certain threshold (RA) [4] and percentile density (PD) [6]. These measures consider only emphysema and treat each parenchyma voxel in the CT image independently, thereby disregarding potentially valuable information such as spatial relations between voxels and patterns at larger scales. The measures are also restricted to a single threshold parameter, which makes them sensitive to scanner calibration and noise in the CT images.

Supervised texture classification in CT, where a classifier is trained on manually annotated regions of interest (ROIs) [7]–[13], uses much more of the information available in the CT images compared to the densitometric measures, and the output of a trained classifier can be used for COPD quantification, e.g., by fusing individual ROI posterior probabilities [11], [13], [14]. However, this approach requires labeled data, which is usually acquired by manual annotation done by human experts. Manual annotation suffers from the same limitations as visual assessment of emphysema in CT images [4], [5], moreover, it is hard to accurately outline regions of emphysema since the appearance of the disease patterns can be subtle and diffuse, especially at early stages of COPD. Further, analysis is limited to current knowledge and experience of the experts, and there can be a bias towards typical cases in the annotated data set. As a consequence, unknown or less typical patterns that are a characteristic part of COPD may not be captured by the trained classifier, and important discriminative information may be disregarded.

In this study, we propose a completely data-driven approach to texture-based analysis of COPD in pulmonary CT images. The main idea is to utilize meta-data that is connected with the CT images to acquire the labels. Hereby, no human intervention is required, and all the above mentioned limitations are handled. Instead, a fully data-driven, and thereby objective, CT image texture-based measure is obtained that can easily be applied to analyze large data sets. Other studies using labels acquired from meta-data, with different features and classification setup, have been published in other areas

of medical image analysis as well, including assessment of structural changes of the breast tissue in digital mammography [15] and detection of tuberculosis in chest radiographs [16].

The proposed approach relies on supervised texture-based classification of ROIs and fusion of individual ROI posterior probabilities similar to [11], [13], [14], but with ROIs and labels obtained in the following way: each CT image is assigned a global label according to PFTs of the scanned subject that are acquired at the same time as the CT image, and ROIs are sampled at random from within the lung fields and labeled with the global label of the CT image. In principle, other meta-data associated with the subject from which the CT image is acquired, such as genetic information and biomarkers from blood samples, could be used when labeling. In this study, PFTs are used, which are the current gold standard for diagnosis of COPD [1]. The obtained texture-based measure is a probability of a subject suffering from COPD based on the evidence in the CT image, and this number reflects COPD severity in two ways. It measures the number of ROIs that show signs of COPD, i.e., how much of the lungs are affected, as well as the individual ROI COPD probabilities, i.e., the confidence about abnormality in individual ROIs.

The performance of the obtained texture-based measure in diagnosis of COPD is compared to the performance of the common densitometric measures RA and PD. A preliminary version of the work presented here appeared in [17]. The analysis is extended to a data set where also mild COPD cases are considered, and several properties of the proposed approach are investigated, including correlation with lung function and the relationship with lung function adjusted for confounders. The reproducibility and robustness to inspiration level is also evaluated.

## II. METHODS

The proposed quantitative measure for COPD relies on texture-based classification of CT ROIs. The ROI classification is done with a  $k$  nearest neighbor ( $k$ NN) classifier using dissimilarity between sets of filter response histograms as distance, and the histograms are based on the filter responses from a rotation invariant, multi-scale Gaussian filter bank [18]. A quantitative measure of the severity of COPD is obtained by fusing the individual ROI posterior probabilities into one posterior probability [14]. This approach has previously been successfully applied on another CT data set using manually labeled ROIs for training [13]. In [13], the same histogram estimation technique was used with two-dimensional versions of a subset of the filters considered in this study, and a quantitative measure of severity was also obtained by fusing ROI posteriors as classified by  $k$ NN, however, the ROI posteriors were estimated using prototype distances.

A segmentation of the lung fields is used in order to steer the sampling of ROIs as well as to decide which voxels contribute to the filter response histograms, and Section II-A describes how this segmentation is obtained. Section II-B describes the sampling procedure. The filter response histograms, or texture descriptors, are described in Section II-C, the ROI classification scheme is described in Section II-D, and the pos-

terior probability fusion for image classification is described in Section II-E.

### A. Segmentation of the lung fields

The lung fields are segmented in CT image  $I$  using a region growing algorithm, which assumes that lung parenchyma is below  $-400$  Hounsfield units (HU). The algorithm automatically detects part of the trachea by searching for a dark cylindrical structure in the top of the image, and the detected trachea is subsequently used to segment the left and right main bronchi. The segmented left and right main bronchi are then used to initiate two region growing procedures that segment the left and right lung field. The final segmented lung fields,  $s(I)$ , are obtained after a post processing step, where erroneously included regions belonging to the esophagus are removed by looking for tube-like structures between the segmented left and right lung fields. This is the same lung segmentation algorithm as is used in [19], which is similar to [20].  $s(I)$  excludes the trachea, the main bronchi, and any structures with CT intensity above  $-400$  HU, which includes part of the vessels, the fissures, and the airway walls. The parameters used in the automatic lung segmentation algorithm were tuned on a different data set than the one used in this study.

### B. Sampling of ROIs

$N_r$ , possibly overlapping, cubic ROIs are sampled at random from within the lung fields of CT image  $I$  according to segmentation  $s(I)$ , and these ROIs represent that image. Only ROIs with centers inside the segmentation are allowed, but parts of an ROI can still be outside the segmentation. These parts are disregarded in the subsequent analysis.

### C. Texture descriptors

In this study, the textural information in a CT image is captured by measuring various texture features in randomly sampled ROIs from that image, and a filtering approach is used for this purpose. A filter bank comprising a total of eight rotational invariant filters based on the Gaussian function and combinations of derivatives of the Gaussian is applied at multiple scales, giving rise to a large number of filtered versions of the CT image. The ROIs in the image are represented by histograms of the filter responses, one for each of the applied filters, and classification is done based on this representation. Steps for obtaining the filter response histograms are given as follows:

1) *Filters*: Eight different measures of local image structure are used as base filters and these are: The Gaussian function

$$G(\mathbf{x}; \sigma) = \frac{1}{(2\pi^{1/2}\sigma)^3} \exp\left(-\frac{\|\mathbf{x}\|_2^2}{2\sigma^2}\right) \quad (1)$$

where  $\sigma$  is the standard deviation, or scale, and  $\mathbf{x} = [x, y, z]^T$  is a voxel; the three eigenvalues of the Hessian matrix

$$\lambda_i(\mathbf{x}; \sigma), \quad i = 1, 2, 3, \quad |\lambda_1| \geq |\lambda_2| \geq |\lambda_3|; \quad (2)$$

gradient magnitude

$$\|\nabla G(\mathbf{x}; \sigma)\|_2 = \sqrt{I_{x,\sigma}^2 + I_{y,\sigma}^2 + I_{z,\sigma}^2} \quad (3)$$

where  $I_{x,\sigma}$  denotes the partial first order derivative of image  $I$  w.r.t.  $x$  at scale  $\sigma$ ; Laplacian of the Gaussian

$$\nabla^2 G(\mathbf{x}; \sigma) = \lambda_1(\mathbf{x}; \sigma) + \lambda_2(\mathbf{x}; \sigma) + \lambda_3(\mathbf{x}; \sigma); \quad (4)$$

Gaussian curvature

$$K(\mathbf{x}; \sigma) = \lambda_1(\mathbf{x}; \sigma)\lambda_2(\mathbf{x}; \sigma)\lambda_3(\mathbf{x}; \sigma); \quad (5)$$

and the Frobenius norm of the Hessian

$$\|H(\mathbf{x}; \sigma)\|_F = \sqrt{\lambda_1(\mathbf{x}; \sigma)^2 + \lambda_2(\mathbf{x}; \sigma)^2 + \lambda_3(\mathbf{x}; \sigma)^2}. \quad (6)$$

Since histograms are used, the ordering of the voxels is disregarded and a classifier can therefore not automatically learn combinations of features such as the Laplacian of the Gaussian from the individual eigenvalues. Combinations of the eigenvalues, i.e., (4), (5), and (6), are therefore explicitly used in the representation.

2) *Normalized convolution*: The filtering is done by normalized convolution [21] with a lung fields segmentation, obtained as described in Section II-A, as binary mask. The equation for normalized convolution is given by

$$I_\sigma = \frac{(S(\mathbf{x})I(\mathbf{x})) * G(\mathbf{x}; \sigma)}{S(\mathbf{x}) * G(\mathbf{x}; \sigma)} \quad (7)$$

where  $*$  denotes convolution and the segmentation  $S = s(I)$  computed from image  $I$  is used as an indicator function, marking whether  $\mathbf{x}$  is a lung parenchyma voxel or not. Derivatives are computed on the Gaussian filtered images using finite differences.

3) *Histogram estimation*: The filter responses are quantized into filter response histograms. The bin widths are derived using adaptive binning [22]. This technique locally adapts the histogram bin widths to the data set at hand such that each bin contains the same mass when computing the histogram of all data. Only voxels in the considered ROI that belong to a lung segmentation  $S$  are used, and the resulting histogram is normalized to sum to one. The number of histogram bins  $N_b$  computed from  $N_s$  voxels is determined according to

$$N_b = \sqrt[3]{N_s}, \quad (8)$$

which is an approximation of the data-based procedure for choosing the bin width presented in [23] that asymptotically minimizes the integrated mean squared error of the histogram estimate of the true density. Using adaptive binning has the potential to further decrease this error.

#### D. ROI Classification

ROI classification is performed using the  $k$  nearest neighbor ( $k$ NN) classifier [24], [25], a nonparametric approach based directly on distances computed between objects.  $k$ NN is the natural classifier of choice when working on a distance representation of the objects and has previously shown to work well for lung tissue classification in CT [8], [13]. The classifier has one free parameter, the number of neighbors  $k$ , that governs the smoothness of the non-linear decision boundary in the feature space. Probabilistic classification outputs for the ROIs

based on the common  $k$ NN posterior probability estimate [25] are used

$$P(\omega_i|\mathbf{x}, I) = \frac{k_{\omega_i}(\mathbf{x})}{k}, \quad \mathbf{x} \in s(I) \quad (9)$$

where  $k_{\omega_i}(\mathbf{x})$  is the number of nearest neighbors of the ROI centered on voxel  $\mathbf{x}$ , from lung segmentation  $s(I)$ , belonging to class  $\omega_i$  out of a total of  $k$  nearest neighbors. The summed histogram dissimilarity is used as ROI distance function in  $k$ NN

$$D(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{N_f} L(f_i(\mathbf{x}), f_i(\mathbf{y})) \quad (10)$$

where  $N_f$  is the number of filter response histograms,  $L(\cdot, \cdot)$  is a histogram dissimilarity measure, and  $f_i(\mathbf{x}) \in \mathbb{R}^{N_b}$  is the  $i$ 'th filter response histogram with  $N_b$  bins estimated from an ROI centered on  $\mathbf{x}$ .

Three histogram dissimilarity measures  $L$  are considered, the L1-norm, the L2-norm, and the earth movers distance (EMD) [26]. The L1-norm and L2-norm are instances of the  $p$ -norm

$$L_p(H, K) = \|H - K\|_p = \left( \sum_{i=1}^{N_b} |H_i - K_i|^p \right)^{1/p} \quad (11)$$

with  $p = 1$  or  $p = 2$  and where  $H, K \in \mathbb{R}^{N_b}$  are histograms each with  $N_b$  bins. The histograms used in this study are normalized to sum to one, and thus  $L_1$  is equivalent to 1 – histogram intersection [26]. EMD can be computed using (11) with  $p = 1$  on cumulative versions of  $H$  and  $K$  when  $H$  and  $K$  are one dimensional, have equal number of bins, and equal mass [27], which is the case in this study. This histogram dissimilarity measure will be denoted by  $L_{\text{EMD}}$ .

#### E. CT Image Classification

The ROI posterior probabilities are combined into an overall subject posterior probability using a static fusion scheme, namely, the mean rule [14]

$$P(\omega_i|I) = \frac{1}{N_r} \sum_{j=1}^{N_r} P(\omega_i|\mathbf{x}_j, I) \quad (12)$$

where  $N_r$  is the number of ROIs that are considered. The average sample posterior probability (12) then provides a measure of the probability that a subject suffers from COPD, based on the CT image. This number reflects both the number of samples that show signs of COPD as well as the probability for the individual ROIs.

## III. EXPERIMENTS AND RESULTS

#### A. Data

Experiments are conducted using low-dose volumetric CT images acquired at full inspiration from current and former smokers enrolled in the Danish Lung Cancer Screening Trial (DLCST) [28] with the following scan parameters: tube voltage 120 kV, exposure 40 mAs, slice thickness 1 mm, and in-plane resolution ranging from 0.72 to 0.78 mm. The subjects were scanned at entry (baseline) and were then subsequently

TABLE I

GROUP CHARACTERISTICS AND LUNG FUNCTION MEASUREMENTS FOR THE HEALTHY AND THE COPD GROUP IN DATA SET  $\mathcal{A}$ . THE NUMBERS REPORTED ARE MEAN VALUES, WITH STANDARD DEVIATION IN PARENTHESES AND RANGE IN SQUARE BRACKETS.

	Healthy	COPD
Age (years)	57 (5) [49-69]	59 (5) [50-71]
Gender (men/women)	172/128	175/125
Height (cm)	174 (9) [150-200]	174 (9) [150-195]
Weight (kg)	77 (14) [50-117]	74 (14) [43-126]
Pack years (years)	35 (12) [19-98]	38 (14) [19-126]
Smoking status (current/former)	222/78	243/57
FEV <sub>1</sub> (L)	3.13 (0.71) [1.26-4.91]	2.62 (0.70) [0.88-4.54]
FEV <sub>1</sub> %pred	1.00 (0.14) [0.58-1.41]	0.85 (0.17) [0.36-1.33]
FEV <sub>1</sub> /FVC	0.76 (0.04) [0.70-0.96]	0.63 (0.06) [0.37-0.70]
GOLD stage (no/I/II/III/IV)	300/0/0/0/0	0/197/93/10/0

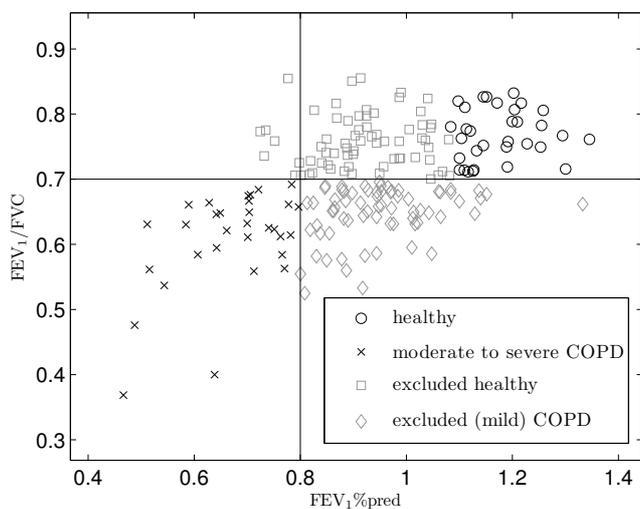


Fig. 1. Construction of  $\mathcal{A}_{\text{train2}}$  from  $\mathcal{A}_{\text{train}}$ . All mild COPD cases and an equal amount of healthy cases with the lowest FEV<sub>1</sub>%pred are excluded from  $\mathcal{A}_{\text{train}}$ .

scanned annually (followup) for four consecutive years. Annual PFTs were also performed along with the CT images, including the forced expiratory volume in one second (FEV<sub>1</sub>) and the forced vital capacity (FVC). Subjects were re-scanned after approximately three months in cases where non-calcified nodules with a diameter of 5 to 15 mm were detected.

We perform experiments on two subsets of the DLCST database that we denote data set  $\mathcal{A}$  and  $\mathcal{B}$ . These data sets are defined in the following way:

**Data set  $\mathcal{A}$ .** Two subject groups are selected at random using the Global Initiative for Chronic Obstructive Lung Disease (GOLD) criteria [1]. The first group consists of 300 subjects without COPD (FEV<sub>1</sub>/FVC  $\geq$  0.7) and the second group consists of subjects with GOLD stage I to III, or mild to severe COPD, (FEV<sub>1</sub>/FVC  $<$  0.7 and FEV<sub>1</sub>%pred  $\geq$  30% where FEV<sub>1</sub>%pred is FEV<sub>1</sub> corrected for age, sex, and height [29]). Baseline CT images from the DLCST database are used. The characteristics of the two groups are reported in Table I.

Data set  $\mathcal{A}$  is randomly divided into three non-overlapping subsets,  $\mathcal{A}_{\text{train}}$ ,  $\mathcal{A}_{\text{validation}}$ , and  $\mathcal{A}_{\text{test}}$  containing 100 healthy and 100 COPD subjects each. A second training set,  $\mathcal{A}_{\text{train2}}$ , is constructed by excluding the mild COPD cases and an equal amount of healthy cases with the lowest FEV<sub>1</sub>%pred from  $\mathcal{A}_{\text{train}}$  as illustrated in Fig. 1. This is to exclude more uncertain cases based on FEV<sub>1</sub>%pred. The number of subjects in  $\mathcal{A}_{\text{train2}}$  is 62, 31 moderate or severe COPD subjects and 31 healthy subjects.

**Data set  $\mathcal{B}$ .** 50 CT image pairs from the DLCST database where both images in a pair are from the same subject that has been re-scanned for a suspicious nodule. All pairs have less than 86 days between the acquisition dates, and we assume progression of COPD to be negligible in this time interval. There is no overlap between the subjects in this data set and the subjects in data set  $\mathcal{A}$ . Data set  $\mathcal{B}$  is used to evaluate the reproducibility and the robustness to inspiration level.

### B. Training and Parameter selection

There are several parameters to select in the proposed classification system and these are listed below together with the possible parameter values considered:

- ROI size  $r \times r \times r$  with  $r = \{21, 31, 41\}$  voxels. The considered ROI sizes are at the resolution of the secondary pulmonary lobule that is approximately 1–2.5 cm in diameter [30], and we assume the lung texture to be homogeneous at this level;
- number of nearest neighbors in the  $k$ NN classifier  $k = \{25, 35, 45\}$ . This was chosen as a trade-off between regularization and speed. The computational complexity of a  $k$  nearest neighbor query in the approximate nearest neighbor library (ANN) [31] is  $\mathcal{O}(kd \log n)$  where  $n$  is the number of prototypes and  $d$  is the dimension. Further, the results did not vary much within the considered range, which is in between 1NN and the usual square root rule  $\sqrt{n} = \sqrt{200 \times 50} = 100\text{NN}$  [32];
- histogram dissimilarity measure  $L = \{L_1, L_2, L_{\text{EMD}}\}$ ;
- the different base filters  $\{(1), (2), (3), (4), (5), (6)\}$  at scales  $\sigma = \{0.6(\sqrt{2})^i\}_{i=0, \dots, 6}$  mm.

The best combination of  $r$ ,  $L$ , and  $k$  is learned using  $\mathcal{A}_{\text{train}}$  or  $\mathcal{A}_{\text{train2}}$  and  $\mathcal{A}_{\text{validation}}$ , and sequential forward feature selection (SFS) [24] is used for determining the optimal histogram subset, from a total of  $N_f = 56$  histograms, for each combination. The objective function in SFS is the area under the receiver operating characteristic (ROC) curve (AUC) computed from the ROIs in  $\mathcal{A}_{\text{validation}}$  classified using  $k$ NN with the ROIs in  $\mathcal{A}_{\text{train}}$  or  $\mathcal{A}_{\text{train2}}$  as prototypes as well as the current values of  $r$ ,  $L$ , and  $k$ .

The posterior probability of COPD in a CT image, i.e.,  $P(\text{COPD}|I)$  estimated using (12), is used as the proposed CT texture-based measure and is denoted  $k$ NN and  $k$ NN2, depending on whether  $\mathcal{A}_{\text{train}}$  or  $\mathcal{A}_{\text{train2}}$  is used.

The number of ROIs sampled per image,  $N_r$ , is fixed to 50, and the number of histogram bins is  $N_b = r$  according to (8). The adaptive histogram binning is computed from  $\mathcal{A}_{\text{train}}$  using a separate set of randomly sampled ROIs, 10 from each image.  $k$ NN classification is performed using the approximate nearest

neighbor (ANN) library [31] with the approximation error set to zero to turn off the approximation part of the algorithm.

### C. Evaluation Strategy

The trained CT texture-based measures,  $kNN$  and  $kNN2$ , are evaluated on  $\mathcal{A}_{\text{test}}$  by ROC analysis on subject level and by Spearman's rank correlation with  $FEV_1\%pred$  and  $FEV_1/FVC$ . The relationship between the obtained CT-based measure and PFTs as well as class label is further investigated using regression-based adjustment for confounders.

We compare the obtained results to the densitometric measures RA and PD. The densitometric measures are computed from the entire lung fields. RA corresponds to the amount of voxels below a given HU threshold relative to the total amount of voxels within the lung fields [4], and a threshold of  $-950$  HU is used in this study. This measure is sometimes referred to as emphysema index or density mask. PD is derived from the CT attenuation histogram as the HU value at a certain percentile [6], and the 15<sup>th</sup> percentile is used in this study. The two densitometric measures are denoted  $RA_{950}$  and  $PD_{15}$ , respectively. Both measures quantify the amount of emphysema present in the lung. However, as for the probabilistic output of the proposed approach, one still needs to choose an operating point to turn the measure into a hard classification of the image. For  $RA_{950}$ , the parameter varied to produce the different points on the ROC-curve is the percentage of voxels below the HU threshold, and for  $PD_{15}$ , the parameter varied is the 15<sup>th</sup> percentile in HU.

### D. COPD diagnosis and quantification

The whole learning framework is applied to data set  $\mathcal{A}$  for COPD diagnosis and quantification using the resulting quantitative measure. The results of the experiment are shown in Fig. 2 and in Table II. The proposed texture-based approach, achieving an AUC of 0.713, is significantly better at discriminating between CT images from healthy subjects and COPD subjects than are the densitometric measures  $PD_{15}$  and  $RA_{950}$  ( $p < 10^{-4}$ ). This is also the case when comparing  $kNN$  and the densitometric measures.  $kNN2$  is significantly correlated with both  $FEV_1\%pred$  and  $FEV_1/FVC$  whereas  $kNN$ ,  $PD_{15}$ , and  $RA_{950}$  are only significantly correlated with  $FEV_1/FVC$ . All four evaluated measures are capable of separating the two subject groups ( $p < 0.05$ ) according to a Wilcoxon rank sum test. The AUC for  $kNN2$  is significantly better than the AUC for  $kNN$ , and  $kNN2$  will therefore be used in the remaining experiments.

Note that the densitometric measures are computed from the full lung fields, and they are therefore based on more information than are the proposed texture-based measure, which is computed from 50 randomly sampled ROIs. The performance of  $PD_{15}$  and  $RA_{950}$  when computed only from the same 50 ROIs as used in  $kNN$  and  $kNN2$ , is slightly worse than when computed from the entire lung fields with  $AUC = 0.586$  and  $AUC = 0.591$ , respectively.

The relationship between the CT-based quantitative measures,  $kNN2$ ,  $RA_{950}$ , and  $PD_{15}$ , and PFTs is investigated while adjusting for confounding factors, namely age, pack

TABLE II

COPD DIAGNOSIS AND QUANTIFICATION RESULTS. AUCs FROM THE ROC ANALYSIS WITH  $p$ -VALUES FOR DIFFERENCE IN AUC WITH  $kNN2$  ACCORDING TO A DELONG, DELONG, AND CLARKE-PEARSON'S TEST [34] SHOWN IN PARENTHESIS. CORRELATION WITH PFTs ACCORDING TO SPEARMAN'S RANK CORRELATION  $\rho$  WITH  $p$ -VALUES OF THE CORRELATION IN PARENTHESIS.

Measure	AUC	$\rho$ $FEV_1\%pred$	$\rho$ $FEV_1/FVC$
$kNN$	0.690 ( $p = 0.045$ )	-0.111 ( $p = 0.116$ )	-0.368 ( $p < 10^{-4}$ )
$kNN2$	0.713 (-)	-0.151 ( $p = 0.033$ )	-0.415 ( $p < 10^{-4}$ )
$RA_{950}$	0.596 ( $p < 10^{-4}$ )	-0.045 ( $p = 0.526$ )	-0.206 ( $p = 0.004$ )
$PD_{15}$	0.598 ( $p < 10^{-4}$ )	0.049 ( $p = 0.492$ )	0.207 ( $p = 0.003$ )

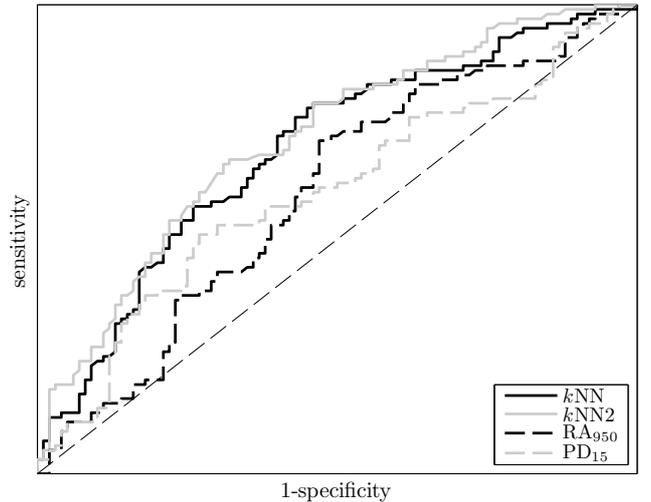


Fig. 2. ROC curves from the experiment. The curves for  $kNN$  and  $kNN2$  are based on (12).

years (one pack year is defined as 20 cigarettes smoked per day for one year), gender, body mass index ( $BMI = \text{weight}/(\text{height}/100)^2$ ), and smoking status [33]. For this purpose the following multiple linear regression model is used

$$PFT = \beta_0 + \beta_1 kNN2 + \beta_2 \text{age} + \beta_3 \text{pack years} + \beta_4 \text{gender} + \beta_5 \text{BMI} + \beta_6 \text{smoking status} + \epsilon \quad (13)$$

where  $\beta_i$  are the free parameters and  $\epsilon$  is a random component. Age and pack years are measured in whole years, gender is binary (0: male, 1: female), and smoking status is binary (0: current smoker, 1: former smoker). The resulting regression parameters estimated on  $\mathcal{A}_{\text{test}}$  are shown in Table III together with associated  $p$ -values. All three CT-based measures significantly explain  $FEV_1/FVC$  according to a t-test whereas none of the CT-based measures significantly explain  $FEV_1\%pred$ . Overall, all six models significantly explain the associated PFT ( $p < 0.05$ ) according to an F-test.

### E. Stability of proposed measure

25% of the volume of the lung segmentation is on average covered by 50 randomly sampled ROIs of size  $41 \times 41 \times 41$  voxels. Theoretically, this could be 35% but some ROIs are overlapping and some ROIs are partly outside the lung segmentation. To inspect whether  $N_r = 50$  is a sufficient



TABLE V  
MEASURES OF REPRODUCIBILITY AND ROBUSTNESS TO INSPIRATION LEVEL, BOTH USING SPEARMAN'S RANK CORRELATION  $\rho(\cdot, \cdot)$ .  $p$ -VALUES OF THE CORRELATIONS ARE SHOWN IN PARENTHESES.

Measure	$\rho(\mathbf{m}_1, \mathbf{m}_2)$	$\rho(\mathbf{m}_2 - \mathbf{m}_1, \mathbf{LV}_{rd})$
$k$ NN2	0.83 ( $p < 10^{-4}$ )	0.71 ( $p < 10^{-4}$ )
RA <sub>910</sub>	0.82 ( $p < 10^{-4}$ )	0.83 ( $p < 10^{-4}$ )
PD <sub>15</sub>	0.81 ( $p < 10^{-4}$ )	-0.83 ( $p < 10^{-4}$ )

from the first image,  $\mathbf{m}_1$ , and measures from the second image,  $\mathbf{m}_2$ , in the 50 pairs of data set  $\mathcal{B}$ . The results are reported in the second column of Table V.  $k$ NN2 is as reproducible as RA<sub>950</sub> and PD<sub>15</sub>.

We use the lung volume (LV) in liters measured from the segmentation as an indicator of the inspiration level. The sensitivity to inspiration level is evaluated by Spearman's rank correlation between signed measurement difference,  $\mathbf{m}_2 - \mathbf{m}_1$ , and the difference in LV divided by the average denoted by  $\mathbf{LV}_{rd}$ . The results are reported in the third column of Table V. Differences in both the densitometric measures and for  $k$ NN2 are significantly correlated with LV difference. However, the correlation with LV difference is lower for  $k$ NN2 compared to the two densitometric measures, indicating slightly lower sensitivity to inspiration level.

#### IV. DISCUSSION AND CONCLUSION

The conducted experiments show that it is possible to train a texture-based classifier to recognize COPD in pulmonary CT images using supervised learning techniques in a fully automatic, data-driven approach without any human intervention. Hereby, all the limitations associated with manual labeling are avoided. The meta-data driven labeling of ROIs, in this study using PFTs, however, has other potential problems. The disease patterns may be localized only in parts of the CT images in subjects with COPD. For instance, paraseptal emphysema is located in the periphery of the lung, centrilobular emphysema is predominantly in the upper lobes, while panlobular emphysema is predominantly in the lower lobes [30]. Randomly sampled ROIs from COPD subjects will therefore likely contain both diseased and healthy tissue where the healthy tissue ROIs still receive the label COPD. The reverse may also be the case in healthy subjects but is expected to be less prominent. The classes in this weakly labeled data set are therefore expected to overlap more compared to classes in manually labeled data where experts have annotated relatively clear examples of the different classes, and this poses a challenging classification problem.

PFTs are insensitive to early stages of COPD [2], lack reproducibility [3], and can be affected by other factors limiting the airflow in the airways than those associated with COPD. Despite these limitations, PFTs were used to obtain labels in this study assuming that it was possible to learn, using supervised learning, the textural COPD patterns in CT that are related to the part of the disease that correlates with PFTs. PFTs are also the current gold standard for COPD diagnosis [1]. The results demonstrate that it is indeed possible to learn the textural patterns associated with PFTs, and it was further

shown that the learned effect still holds when adjusting for confounding factors such as age and smoking history.

Including mild COPD cases in the analysis deteriorates performance in terms of AUC compared to [17] that excluded these cases. There are three possibilities that could cause this result. First, the disease may be picked up by PFTs before textural changes appear in the lung tissue. Including mild COPD, as diagnosed by PFTs, may in this case add data that the proposed approach cannot distinguish. Secondly, PFTs may pick up the disease at the same time as textural changes appear, in which case including mild COPD simply produces a harder classification problem since the signal is weaker in this data. Thirdly, textural changes may appear before PFTs pick up the disease. In this case, data with more uncertain labels is added when considering mild COPD cases. We hypothesize that the third possibility is the case, since PFTs measure the respiratory system as a whole as opposed to texture that measures local changes in the lung tissue. Further, it has been estimated that 30% of the lung must be destroyed by emphysema before pulmonary function abnormalities become evident [2] whereas loss of lung tissue is apparent earlier in a CT image.

A two-class problem was defined by the two subject groups, healthy (no COPD according to the GOLD criteria [1]), and COPD (mild to severe COPD or moderate to severe COPD according to the GOLD criteria [1]). However, other possibilities exist, both on the type of problem to consider and on the type of meta-data to use for group definitions. One possibility would be to consider several or all of the four GOLD stages [1] as separate groups, which is similar in spirit to [35], for assessing GOLD stage or COPD severity. However, regression may be more suitable for this purpose. The proposed approach could also be used to gain a better understanding of which textural patterns in the CT images that are related to, e.g., different genotypes or markers from blood samples by using genetic information or blood biomarkers to define groups and apply the whole learning framework. This may be expanded to further analyze how these patterns evolve over time in longitudinal data.

The proposed approach is directly applicable in situations where a large data set is available and an objective quantitative measure for specifically that data is needed, as long as labeled training data can be obtained. This could be in pharmaceutical studies or in clinical research, for example. However, the trained method is tied to data with the same or similar characteristics. Application in clinical practice is less straightforward due to large variety in data, mainly caused by different scanning protocols. Implementation in a workstation aimed at clinical use would at least require a reliable segmentation of the lung fields and an agreed upon set of base filters. Lung segmentation algorithms are already implemented in several commercial workstations. Based on the results of in this study and on previous results [13], it could be the following base filters:  $\|\nabla(\mathbf{x}; \sigma)\|_2$ ,  $\nabla^2(\mathbf{x}; \sigma)$ , and  $G(\mathbf{x}; \sigma)$ . The specific scale at which the different base filters are applied can be tuned according to the scanner settings, either automatically, e.g., using SFS, or by setting them according to scan parameters such as the used reconstruction kernel and radiation dose.

How the prototypes are selected for the  $k$ NN classifier influences the performance of the resulting measure, as demonstrated in the results of Section III-D. Systematically excluding mild COPD cases and healthy cases with a low FEV<sub>1</sub>%pred from the training set resulted in a significantly larger AUC. This observation probably has a big impact in a clinical setting where the data variability is large, and proper selection of prototypes should be further investigated.

Classifiers were trained at the ROI level without including knowledge about the subsequent fusion of the ROI posterior probabilities using (12). The rationale is that we would like to capture the local texture information and use this for quantification. Although the proposed approach works well as illustrated in the results, it remains an open research question whether training locally followed by posterior fusion is the best approach when the final goal is quantification at CT image level. An alternative approach would be to take CT image level information into account during training, e.g., by adapting the objective function for SFS to use (12) instead of (9).

COPD comprises two main components, small airway disease and emphysema [1]. The proposed approach measures parenchymal texture and therefore mainly targets emphysema. However, small airway disease is included to some extent since the lung fields segmentation includes the small airways and since the labels are obtained from PFTs, which are affected by both components. The airway information could be targeted more explicitly by combining the output of the proposed approach with measurements computed directly on the segmented airway tree, e.g., [36], which may provide a more accurate measure of COPD.

In [13] it was shown that rotation invariant local binary patterns (LBPs) and a rotation invariant filter bank based on Gaussian derivatives achieved similar performance on an emphysema quantification task in 2D CT images. We therefore chose to use a superset of the filter bank in this study as this texture descriptor is directly extendable to 3D. However, as mentioned in [13], some approximate 3D extensions of rotation invariant LBP do exist.

In conclusion, we have proposed a fully automatic, data-driven approach for texture-based quantitative analysis of COPD in pulmonary CT. The obtained texture-based measure demonstrates superior performance in discriminating between subjects with and without COPD compared to the common densitometric measures RA and PD, with an AUC of 0.713 compared to 0.596 and 0.598, respectively. Further, the proposed approach is as reproducible as the densitometric measures, and there are indications that it correlates better with lung function and is less sensitive to inspiration level – a major source of variability in computerized quantitative CT measures. Since the method does not rely on labels annotated by human experts, the resulting CT image-based measure is objective and can easily be applied to analyze large data sets. Overall, the proposed approach results in a robust and objective measure that can facilitate better computerized quantification of COPD in pulmonary CT.

## REFERENCES

- [1] K. F. Rabe, S. Hurd, A. Anzueto, P. J. Barnes, S. A. Buist, P. Calverley, Y. Fukuchi, C. Jenkins, R. Rodriguez-Roisin, C. van Weel, and J. Zielinski, "Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary," *Am. J. Respir. Crit. Care Med.*, vol. 176, no. 6, pp. 532–555, Sep 2007.
- [2] J. W. Gurney, K. K. Jones, R. A. Robbins, G. L. Gossman, K. J. Nelson, D. Daughton, J. R. Spurzem, and S. I. Rennard, "Regional distribution of emphysema: correlation of high-resolution CT with pulmonary function tests in unselected smokers," *Radiology*, vol. 183, no. 2, pp. 457–463, 1992.
- [3] A. Dirksen, N. H. Holstein-Rathlou, F. Madsen, L. T. Skovgaard, C. S. Ulrik, T. Heckscher, and A. Kok-Jensen, "Long-range correlations of serial FEV1 measurements in emphysematous patients and normal subjects," *J. Appl. Physiol.*, vol. 85, no. 1, pp. 259–265, Jul 1998.
- [4] N. L. Müller, C. A. Staples, R. R. Miller, and R. T. Abboud, "'Density mask'. An objective method to quantitate emphysema using computed tomography," *Chest*, vol. 94, no. 4, pp. 782–787, Oct 1988.
- [5] A. A. Bankier, V. D. Maertelaer, C. Keyzer, and P. A. Gevenois, "Pulmonary emphysema: subjective visual grading versus objective quantification with macroscopic morphometry and thin-section CT densitometry," *Radiology*, vol. 211, no. 3, pp. 851–858, Jun 1999.
- [6] G. A. Gould, W. MacNee, A. McLean, P. M. Warren, A. Redpath, J. J. Best, D. Lamb, and D. C. Flenley, "CT measurements of lung density in life can quantitate distal airspace enlargement—an essential defining feature of human emphysema," *Am. Rev. Respir. Dis.*, vol. 137, no. 2, pp. 380–392, Feb 1988.
- [7] F. Chabat, G.-Z. Yang, and D. M. Hansell, "Obstructive lung diseases: texture classification for differentiation at CT," *Radiology*, vol. 228, no. 3, pp. 871–877, Sep 2003.
- [8] I. C. Sluimer, P. F. van Waes, M. A. Viergever, and B. van Ginneken, "Computer-aided diagnosis in high resolution CT of the lungs," *Med. Phys.*, vol. 30, no. 12, pp. 3081–3090, Dec 2003.
- [9] R. Uppaluri, E. A. Hoffman, M. Sonka, P. G. Hartley, G. W. Hunninghake, and G. McLennan, "Computer recognition of regional lung disease patterns," *Am. J. Respir. Crit. Care Med.*, vol. 160, no. 2, pp. 648–654, Aug 1999.
- [10] Y. Xu, M. Sonka, G. McLennan, J. Guo, and E. A. Hoffman, "MDCT-based 3-D texture classification of emphysema and early smoking related lung pathologies," *IEEE Trans. Med. Imag.*, vol. 25, no. 4, pp. 464–475, Apr 2006.
- [11] Y. S. Park, J. B. Seo, N. Kim, E. J. Chae, Y. M. Oh, S. D. Lee, Y. Lee, and S.-H. Kang, "Texture-based quantification of pulmonary emphysema on high-resolution computed tomography: comparison with density-based quantification and correlation with pulmonary function test," *Invest. Radiol.*, vol. 43, no. 6, pp. 395–402, June 2008.
- [12] M. Prasad, A. Sowmya, and P. Wilson, "Multi-level classification of emphysema in HRCT lung images," *Pattern Anal. Appl.*, vol. 12, no. 1, pp. 9–20, Jan 2009.
- [13] L. Sørensen, S. B. Shaker, and M. de Bruijne, "Quantitative analysis of pulmonary emphysema using local binary patterns," *IEEE Trans. Med. Imag.*, vol. 29, no. 2, pp. 559–569, Feb. 2010.
- [14] M. Loog and B. Van Ginneken, "Static posterior probability fusion for signal detection: applications in the detection of interstitial diseases in chest radiographs," in *International Conference on Pattern Recognition*. IEEE Computer Society, 2004, pp. 644–647.
- [15] J. Raundahl, M. Loog, P. Pettersen, L. B. Tanko, and M. Nielsen, "Automated effect-specific mammographic pattern measures," *IEEE Trans. Med. Imag.*, vol. 27, no. 8, pp. 1054–1060, Aug. 2008.
- [16] Y. Arzhaeva, L. Hogeweg, P. A. de Jong, M. A. Viergever, and B. van Ginneken, "Global and local multi-valued dissimilarity-based classification: application to computer-aided detection of tuberculosis," in *Lecture Notes in Computer Science, G.-Z. Yang, D. J. Hawkes, D. Rueckert, J. A. Noble, and C. J. Taylor, Eds.*, vol. 5762. Springer, 2009, pp. 724–731.
- [17] L. Sørensen, P. Lo, H. Ashraf, J. Sporring, M. Nielsen, and M. de Bruijne, "Learning COPD sensitive filters in pulmonary CT," in *Medical Image Computing and Computer-Assisted Intervention*, ser. Lecture Notes in Computer Science, G.-Z. Yang, D. J. Hawkes, D. Rueckert, J. A. Noble, and C. J. Taylor, Eds., vol. 3. Springer, 2009, pp. 699–706.
- [18] B. M. ter Haar Romeny, *Gaussian Scale-Space Theory*. Dordrecht: Kluwer Academic Publishers, 1997, ch. Applications of scale-space theory, pp. 3–19.

- [19] P. Lo, J. Sporning, H. Ashraf, J. J. Pedersen, and M. de Bruijne, "Vessel-guided airway tree segmentation: A voxel classification approach," *Med. Image Anal.*, vol. 14, pp. 527–538, Aug 2010.
- [20] S. Hu, E. A. Hoffman, and J. M. Reinhardt, "Automatic lung segmentation for accurate quantitation of volumetric X-ray CT images," *IEEE Trans. Med. Imag.*, vol. 20, no. 6, pp. 490–498, 2001.
- [21] H. Knutsson and C.-F. Westin, "Normalized and differential convolution: Methods for interpolation and filtering of incomplete and uncertain data," in *Computer Vision and Pattern Recognition*. IEEE Computer Society, 1993, pp. 515–523.
- [22] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern Recognit.*, vol. 29, no. 1, pp. 51–59, 1996.
- [23] D. W. Scott, "On optimal and data-based histograms," *Biometrika*, vol. 66, no. 3, pp. 605–610, Dec 1979.
- [24] A. Jain, R. Duin, and J. Mao, "Statistical pattern recognition: a review," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 1, pp. 4–37, Jan. 2000.
- [25] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification (2nd Edition)*. Wiley-Interscience, 2000.
- [26] Y. Rubner, C. Tomasi, and L. J. Guibas, "The earth mover's distance as a metric for image retrieval," *Int. J. Comput. Vision*, vol. 40, no. 2, pp. 99–121, 2000.
- [27] E. Levina and P. Bickel, "The earth mover's distance is the mallows distance: some insights from statistics," in *International Conference on Computer Vision*. IEEE Computer Society, 2001, pp. 251–256.
- [28] J. H. Pedersen, H. Ashraf, A. Dirksen, K. Bach, H. Hansen, P. Toennesen, H. Thorsen, J. Brodersen, B. G. Skov, M. Døssing, J. Mortensen, K. Richter, P. Clementsen, and N. Seersholm, "The Danish randomized lung cancer CT screening trial—overall design and results of the prevalence round," *J. Thorac. Oncol.*, vol. 4, no. 5, pp. 608–614, May 2009.
- [29] P. Quanjer, G. Tammeling, J. Cotes, O. Pedersen, R. Peslin, and J. Yernault, "Lung volumes and forced ventilatory flows. Report working party standardization of lung function tests, European Community for Steel and Coal. Official statement of the European Respiratory Society," *Eur Respir J Suppl*, vol. 16, pp. 5–40, 1993.
- [30] W. R. Webb, N. Müller, and D. Naidich, *High-Resolution CT of the Lung, Third Edition*, J.-R. John, Ed. Lippincott Williams & Wilkins, 2001.
- [31] S. Arya, D. M. Mount, N. S. Netanyahu, R. Silverman, and A. Y. Wu, "An optimal algorithm for approximate nearest neighbor searching in fixed dimensions," *J. ACM*, vol. 45, no. 6, pp. 891–923, 1998.
- [32] J. Kittler and F. M. Alkoot, "Moderating k-NN classifiers," *Pattern Analysis and Applications*, vol. 5, no. 3, pp. 326–332, 2002.
- [33] H. Ashraf, P. Lo, S. B. Shaker, M. de Bruijne, A. Dirksen, P. Tønnenesen, M. Dahlbäck, and J. H. Pedersen, "Short-term effect of changes in smoking behaviour on emphysema quantification by CT," *Thorax*, vol. 66, no. 1, pp. 55–60, Jan 2011.
- [34] E. R. DeLong, D. M. DeLong, and D. L. Clarke-Pearson, "Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach," *Biometrics*, vol. 44, no. 3, pp. 837–845, Sep 1988.
- [35] K. Murphy, B. van Ginneken, E. van Rikxoort, B. de Hoop, M. Prokop, P. Lo, M. de Bruijne, and J. Pluim, "Obstructive pulmonary function: patient classification using 3D registration of inspiration and expiration CT images," in *The Second International Workshop on Pulmonary Image Analysis*, M. Brown, M. de Bruijne, B. van Ginneken, A. Kiraly, J.-M. Kuhnigk, C. Lorenz, J. McClelland, K. Mori, A. Reeves, and J. Reinhardt, Eds. CreateSpace, 2009, pp. 37–47.
- [36] J. Petersen, P. Lo, M. Nielsen, G. Edula, H. Ashraf, A. Dirksen, and M. de Bruijne, "Quantitative analysis of airway abnormalities in CT," in *Medical Imaging: Computer-Aided Diagnosis*, N. Karssemeijer and R. M. Summers, Eds., vol. 7624. SPIE, 2010, p. 6241S.