

# Automated Interpretation of Pulmonary Function Tests in Adults with Respiratory Complaints

Marko Topalovic<sup>a</sup> Stefan Laval<sup>b</sup> Jean-Marie Aerts<sup>c</sup> Thierry Troosters<sup>a,d</sup>  
Marc Decramer<sup>a</sup> Wim Janssens<sup>a</sup> on behalf of the Belgian Pulmonary Function  
Study investigators

<sup>a</sup>Laboratory of Respiratory Diseases, Department of Clinical and Experimental Medicine, University Hospital Leuven, KU Leuven, Leuven, Belgium; <sup>b</sup>Sana Kliniken Lübeck, Medical Department, Pneumology, Lübeck, Germany; <sup>c</sup>Division of Animal and Human Health Engineering, Department of Biosystems, KU Leuven, and <sup>d</sup>Department of Rehabilitation Sciences, Faculty of Kinesiology and Rehabilitation Sciences, KU Leuven, Leuven, Belgium

## Keywords

Pulmonary function tests · Decision tree · Interpretative strategies · Lung diseases

## Abstract

**Background:** The use of pulmonary function tests is primarily based on expert opinion and international guidelines. Current interpretation strategies are using predefined cut-offs for the description of a typical pattern. **Objectives:** We aimed to explore the predicted disease outcome based on the American Thoracic Society/European Respiratory Society (ATS/ERS) interpreting strategy. Subsequently, we investigated whether an unbiased machine learning framework integrating lung function with clinical variables may provide alternative decision trees resulting in a more accurate diagnosis. **Methods:** Our study included data from 968 subjects admitted for the first time to a pulmonary practice. The final clinical diagnosis was based on the combination of complete pulmonary function with the investigations that were decided at the physician's discretion. Clinical diagnoses were separated into 10 different groups and validated by an expert panel. **Results:** The ATS/ERS algorithm resulted in a correct diagnostic label in 38% of the subjects. Chronic ob-

structive pulmonary disease (COPD) was detected with an acceptable accuracy (74%), whereas all other diseases were poorly identified. The new data-based decision tree improved the general accuracy to 68% after 10-fold cross-validation when detecting the most common lung diseases, with a significantly higher positive predictive value and sensitivity for COPD, asthma, interstitial lung disease, and neuromuscular disorder (83/78, 66/82, 52/59, and 100/54%, respectively). **Conclusions:** Our data show that the current algorithms for lung function interpretation can be improved by a computer-based choice of lung function and clinical variables and their decision-making thresholds.

© 2017 S. Karger AG, Basel

## Introduction

Pulmonary function tests (PFTs) are the prime tool of respiratory physicians. In clinical medicine, PFTs are often used to evaluate respiratory symptoms, to diagnose diseases, and to assess functionality and preoperative risk [1]. After taking the patient's history, the PFT is often combined with blood analysis, lung imaging, and more specific tests (such as skin prick tests, bronchial challenge,

exercise testing, bronchoscopy with biopsies, bronchoalveolar lavage, etc. [2–4]) to come to a final diagnosis [5]. When considering the value of PFTs alone, the Belgian Pulmonary Function Study (BPFS) clearly showed that spirometry, resistance, lung volume, and diffusing capacity significantly and independently contributed to the diagnostic workup [6]. When combined with clinical history, PFTs had an accuracy of 77% in predicting the diagnosis, which may indicate that these tests (when correctly interpreted) have a high potential for an appropriate diagnostic labeling.

The use of PFTs is primarily based on expert opinion and international guidelines predominantly dealing with asthma, chronic obstructive pulmonary disease (COPD), and lung fibrosis [7, 8]. In 2005, an American Thoracic Society/European Respiratory Society (ATS/ERS) task force designed a simplified algorithm to assess lung function in clinical practice [9]. It involves the recognition of typical patterns with strictly defined cutoffs for abnormality (obstructive, restrictive, or mixed pattern) and grades severity over time. Various algorithms have incorporated this strategy, and most PFT equipment has now built-in software that can generate a preliminary interpretation [10–12]. Unfortunately, these algorithms are not well established; they remain rather descriptive, do not deal with atypical patterns, and do not rely on clinical characteristics for diagnostic suggestion. In daily practice, these shortcomings are tackled by an expert clinical reading, which may explain why none of these automated protocols have found their way to clinical routine. Interestingly, the opposite has happened with automated protocols for electrocardiography (ECG) recordings [13, 14].

In this study, we aimed to explore whether the interpretation of PFTs for respiratory disease labeling can be automated. By developing a computer algorithm based on the ATS/ERS interpreting strategy, we checked the predicted disease outcome in a large population. Subsequently, we investigated whether a machine learning framework may provide new decision algorithms that result in a more accurate diagnosis.

## Subjects and Methods

### Study Population

The study included the data of 968 subjects from the BPFS, a prospective cohort study that enrolled a clinical population-based sample of all successive undiagnosed patients admitted for the first time to one of the 33 participating Belgian hospitals due to respiratory symptoms [6]. The study was performed in the periods from June 6 to September 12, 2011, and from January 16 to June 12, 2012. Briefly, all enrolled subjects were Caucasians between 18 and

75 years old who had performed a complete PFT at cohort entry (including post-bronchodilator spirometry, whole-body plethysmography for lung volume and airway resistance, and diffusing capacity). To establish a respiratory disease diagnosis, all necessary additional tests including imaging, ECG, and other PFTs were performed at the physician's discretion. The final diagnosis for each subject was subsequently validated by Belgian local focus groups (groups of 20–25 pulmonologists) who jointly evaluated all tests outcomes. The study population's baseline characteristics are shown in Table 1, covering healthy controls as well as a wide range of respiratory diseases that may present with a disturbed PFT with a specific pattern: asthma, COPD, other obstructive diseases (including bronchiolitis, bronchiectasis, and cystic fibrosis), upper airway obstruction, obesity, interstitial lung disease (ILD), systemic sclerosis or vascular disease, cardiac failure, and hyperventilation. Because of similar lung function disturbances, neuromuscular disorder (NMD) combined patients with chest wall or pleural disease, lung resection, or true neuromuscular disease. Others diagnoses (lung cancer, rhinosinusitis, etc.) were excluded from the analysis. The study protocol was approved by the Ethics Committee of the University Hospital in Leuven, which acted as the leading Ethics Committee, and by all the Ethics Committees of the participating hospitals which acted as subsidiary Ethics Committees. All included patients provided informed consent. The BPFS design can be found on [www.clinicaltrials.gov](http://www.clinicaltrials.gov) (NCT01297881).

### Pulmonary Function Tests

All PFTs were performed according to the ATS/ERS criteria [15] using standardized equipment (Masterscreen Jaeger; CareFusion, Germany). Spirometry data are post-bronchodilator measures and are expressed, along with plethysmography measurements of airway resistance and lung volumes, as percent predicted of normal reference values [16, 17]. Diffusing capacity for carbon monoxide ( $D_{LCO}$ ) was measured by the single-breath gas transfer method, using a standardized carbon monoxide and helium gas mixture, and expressed as percent predicted of reference values [18].

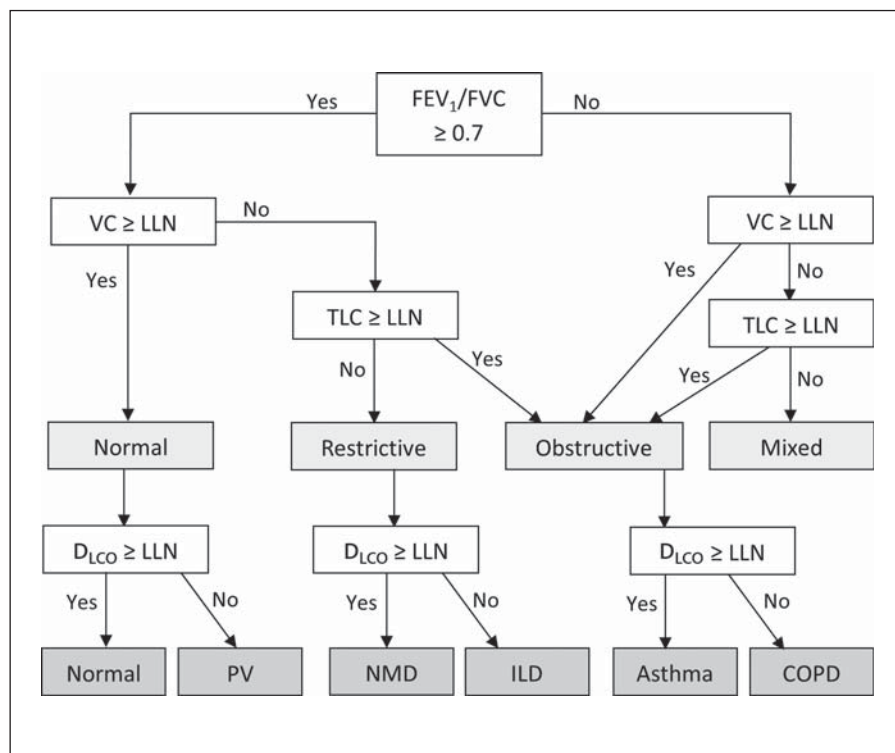
### PFT Interpretation Algorithm

To interpret the PFTs, the international guidelines of the ATS and the ERS were applied [9]. An important modification was made for the cutoff of  $FEV_1/VC$ , where we applied a post-bronchodilator  $FEV_1/FVC$  fixed ratio of 0.7, in accordance with the GOLD and our local clinical recommendations for the diagnosis of obstructive airways disease [8] (Fig. 1).

### Computer Algorithm and Validation

The development of a computer algorithm to automatically interpret all PFTs was performed in MATLAB 8.3 (The MathWorks, Natick, MA, USA). Using the Statistics and Machine Learning Toolbox, a new decision tree was developed based on lung function data (included measurements:  $FEV_1$ , FVC,  $FEV_1/FVC$ , PEF, FEF25, FEF50, FEF75, FEF25–75,  $R_{AW}$ ,  $sG_{AW}$ , VC, RV, TGV, TLC,  $D_{LCO}$ ,  $K_{CO}$ ), combined with the patient characteristics age, pack-years, CAT score, gender, and BMI. The best split criterion was determined using maximal deviance reduction [19]. To avoid “overfitting” of the model and to get a better sense of the predictive accuracy of a decision tree, a stratified 10-fold cross-validation was performed [20]. Briefly, the training data were randomly split into 10 equal segments, and 10 new trees were trained on 9 segments

**Fig. 1.** Modified ATS/ERS algorithm for assessing pulmonary function tests in clinical practice. COPD, chronic obstructive pulmonary disease;  $D_{LCO}$ , diffusing capacity for carbon monoxide;  $FEV_1$ , forced expiratory volume in 1 s; FVC, forced vital capacity; ILD, interstitial lung disease; LLN, lower limit of normal; NMD, neuromuscular disorder; PV, pulmonary vascular disorders; TLC, total lung capacity; VC, vital capacity.

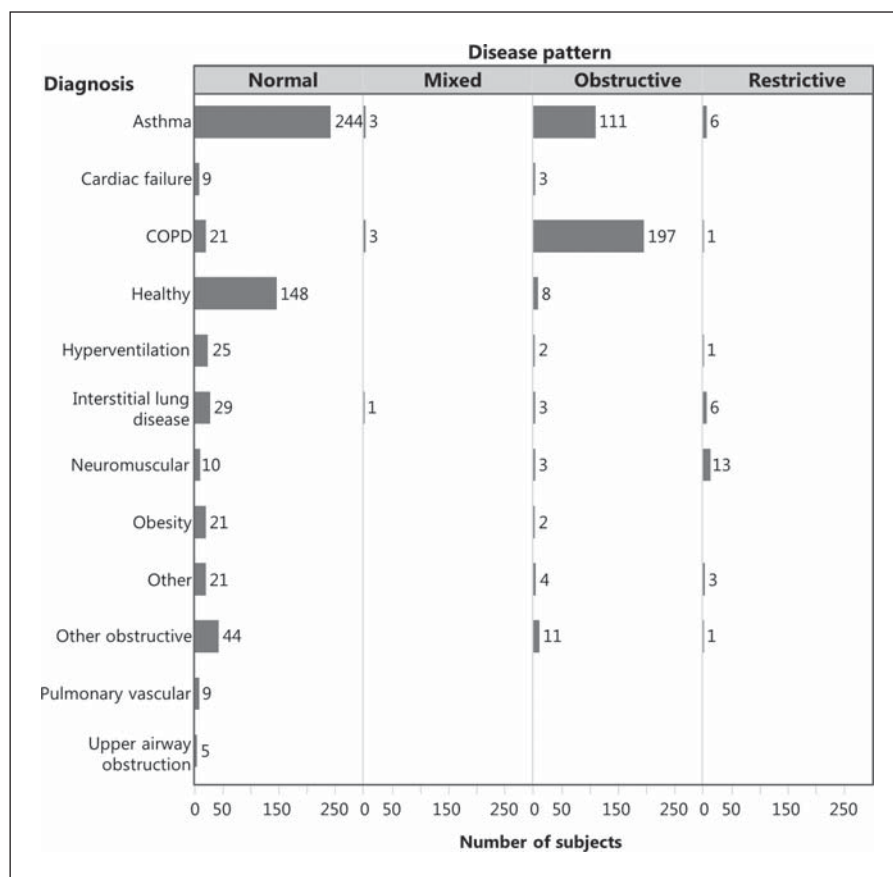


**Table 1.** Population characteristics

	Healthy	Asthma	COPD	Other obstructive	Heart failure	ILD	PV	NMD	HV	Obesity
Subjects, <i>n</i>	156	364	222	56	12	39	9	26	28	23
Gender, M/F	81/75	164/200	127/95	28/28	6/6	23/16	5/4	18/8	11/17	14/9
Age, years	54 (42–64)	50 (34–60)	62 (53–69)	57 (47–67)	72 (59–78)	63 (52–72)	64 (57–71)	54 (41–65)	45 (35–52)	52 (44–64)
$FEV_1$ , %pred.	103 (92–114)	92 (80–102)	64 (51–80)	89 (78–100)	91 (78–107)	85 (72–107)	98 (87–110)	67 (59–81)	101 (93–108)	86 (74–100)
FVC, %pred.	107 (97–115)	101 (91–112)	88 (73–105)	96 (87–110)	95 (80–113)	86 (74–104)	99 (87–118)	70 (56–82)	109 (102–115)	86 (77–100)
$FEV_1/FVC$ , %	80 (75–83)	76 (68–81)	60 (51–66)	74 (70–80)	74 (72–79)	80 (75–85)	77 (73–83)	81 (76–83)	82 (78–84)	81 (78–83)
TLC, %pred.	105 (96–112)	104 (93–115)	112 (99–124)	102 (93–111)	85 (81–99)	85 (75–92)	97 (83–109)	79 (67–85)	108 (97–113)	82 (78–91)
$D_{LCO}$ , %pred.	86 (77–99)	86 (78–95)	59 (48–70)	77 (66–91)	79 (60–88)	54 (41–63)	69 (47–86)	67 (50–80)	82 (72–90)	83 (75–95)
$K_{CO}$ , %pred.	94 (82–104)	98 (86–110)	74 (58–91)	92 (83–107)	101 (77–114)	74 (67–83)	78 (69–106)	107 (95–122)	90 (79–97)	110 (95–123)

Values of other disease diagnoses (lung cancer, NTE problems, etc.) and upper airway obstruction are not reported ( $n = 33$ ). Values are medians and IQR unless indicated otherwise. COPD, chronic obstructive pulmonary disease;  $D_{LCO}$ , diffusing capacity for carbon monoxide;  $FEV_1$ , forced expiratory volume in 1 s; FVC, forced vital capacity; HV, hyperventilation; ILD, interstitial lung disease;  $K_{CO}$ , transfer coefficient for carbon monoxide; NMD, neuromuscular disorder; PV, pulmonary vascular disorders; TLC, total lung capacity.

**Fig. 2.** Distribution of each lung disease within each pattern of abnormality according to the ATS/ERS decision tree. ATS, American Thoracic Society; COPD, chronic obstructive pulmonary disease; ERS, European Respiratory Society.



and validated on the data from the segment not included in training. This method gives a better estimation of the predictive accuracy of the produced decision tree, since it tests and improves new trees on new data.

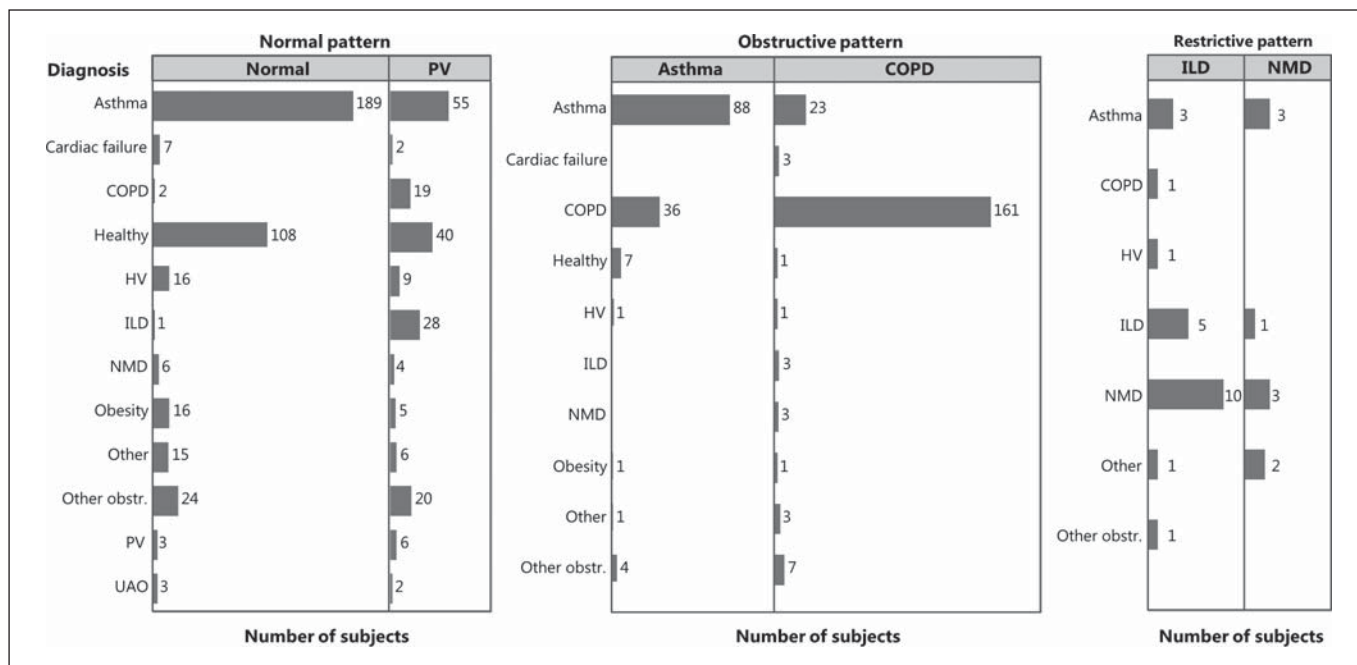
## Results

### *Abnormalities and Diagnosis*

Applying the ATS/ERS algorithm revealed that in a real clinical population sample, the most prevalent lung function pattern was the healthy one (60%) followed by an obstructive pattern (36%), whereas restrictive and mixed patterns were very uncommon (4% together). Only 5% of the truly healthy subjects had an  $FEV_1/FVC$  ratio  $<0.7$ , yet only 25% of all subjects with a normal pattern were truly healthy (Fig. 2). As expected, the majority of healthy patterns were found in patients with asthma, who are known to have normal pulmonary function in stable conditions. Of the 222 COPD patients, 197 had an obstructive pattern according to the modified ATS/ERS rules; 25 had an  $FEV_1/FVC$  ratio  $>0.7$  and may have been labeled as COPD pa-

tients by the expert panel because of hyperinflation, high resistance, low  $D_{LCO}$ , or emphysema on CT. An obstructive pattern was also retrieved in 30% of the asthma subjects. A purely restrictive pattern was rarely seen in COPD (only 1 patient), and although some patients with ILD, NMD, or obesity presented within this restrictive subgroup, the majority were found under the healthy label. When further applying the ATS/ERS algorithm, the lower limit of normal (LLN) for  $D_{LCO}$  was used as a cutoff to split the predicted disease patterns (Fig. 3). It helped in the differentiation of asthma from COPD within the obstructive group, but was not selective enough to identify pulmonary vascular disease in the group with normal pattern or ILD in the group with restrictive pattern. Despite the fact that the largest fraction of patients with real ILD and NMD were found with a normal lung function pattern, once categorized within the restrictive group,  $D_{LCO}$  was able to correctly classify 5 of the 6 ILD patients. Unfortunately,  $D_{LCO}$  was not able to identify NMD within the restrictive labels, as 10 patients out of 13 were misdiagnosed with ILD. In the complete cohort, 38% of subjects were correctly classified. If restricting the cohort to diseases that





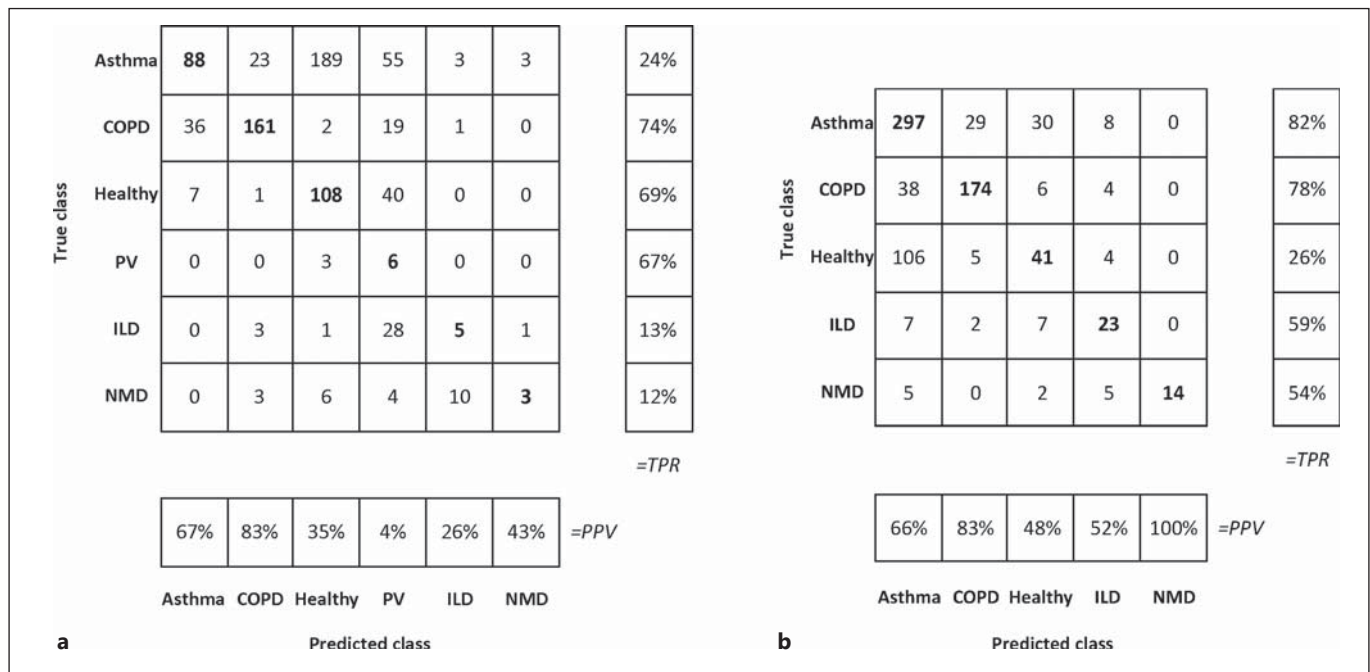
**Fig. 3.** Distribution of each respiratory disease categorized by the ATS/ERS algorithm into diagnostic classes. COPD, chronic obstructive pulmonary disease; HV, hyperventilation; ILD, interstitial lung disease; NMD, neuromuscular disorder; PV, pulmonary vascular disorders; UAO, upper airway obstruction.

can be labeled by the ATS/ERS algorithm (healthy, asthma, COPD, PV, ILD, and NMD;  $n = 810$ ), accuracy increased to 46%. It resulted in 24% sensitivity and 94% specificity for the diagnosis of asthma, 73% sensitivity and 94% specificity for COPD, and 13% sensitivity and 98% specificity for ILD. A detailed comparison of correct and incorrect classifications is presented in Figure 4a.

#### *The Decision Tree Suggests the Final Diagnosis*

As the computer development of new decision trees is very much affected by disease prevalence, we only included data from the most common lung diseases (asthma, COPD, ILD, NMD) and added the healthy individuals. PV was only present in 9 cases (<1%) and was therefore not included in the analysis. Applying the modified ATS/ERS tree to this group ( $n = 801$ ) of most common diseases resulted in a similar accuracy of 46% (365/801) for correct diagnosis. We then developed a computer-based algorithm to define upfront a 100% specific lung function pattern for NMD (Fig. 5a), as its prevalence in the BPFs was still too low to accurately discriminate within the entire population. Fourteen out of 26 NMD subjects were a priori selected as having a unique lung function pattern of NMD. Next, all lung function data and a selected set of clinical parameters (age, pack-years, CAT score, gender,

and BMI) of the remaining population (asthma, COPD, ILD, and healthy subjects;  $n = 784$ ) were subjected to a machine learning framework to develop a decision algorithm. This tree, visualized in Figure 5b, resulted in 74% accuracy on the training data, which decreased to 68% after 10-fold cross-validation. Most interestingly, the decision tree started with a diffusing capacity cutoff of 70% predicted as first discriminator, followed by an FEV<sub>1</sub>/FVC ratio cutoff around 70% predicted. On the lower levels of the tree, pack-years, age, PEF, and TLC were all found to offer significant discriminative power to the tree. A confusion matrix on the decision algorithm that combined the a priori identification of NMD with the tree for the more prevalent disease on all 801 subjects is shown in Figure 4b. It demonstrates that the new algorithm was able to recognize COPD with a high positive predictive value (PPV = 83%) and sensitivity (true positive rate [TPR] = 78%). The proposed algorithm was much stronger to predict the presence of asthma (PPV = 66%, TPR = 82%) and ILD (PPV = 52%, TPR = 59%). For the NMD pattern, a TPR of 54% and a PPV of 100% were reached. Obviously, the decision rules were specific enough, crossing the 90% specificity for every disease. Defining a clear healthy pattern gave the least favorable results, as it is often confused with non-obstructed asthma.



**Fig. 4. a** Confusion matrix with all correctly and incorrectly classified subjects using the ATS/ERS algorithm. There are 6 labels ( $n = 810$ ) correctly classified in 46% of cases. **b** Confusion matrix of the newly developed decision tree after 10-fold cross-validation (accuracy 68%). ATS, American Thoracic Society; COPD, chronic obstructive pulmonary disease; ERS, European Respiratory Society; ILD, interstitial lung disease; NMD, neuromuscular disorder; PPV, positive predictive value; PV, pulmonary vascular disorders; TPR, true positive rate = sensitivity.

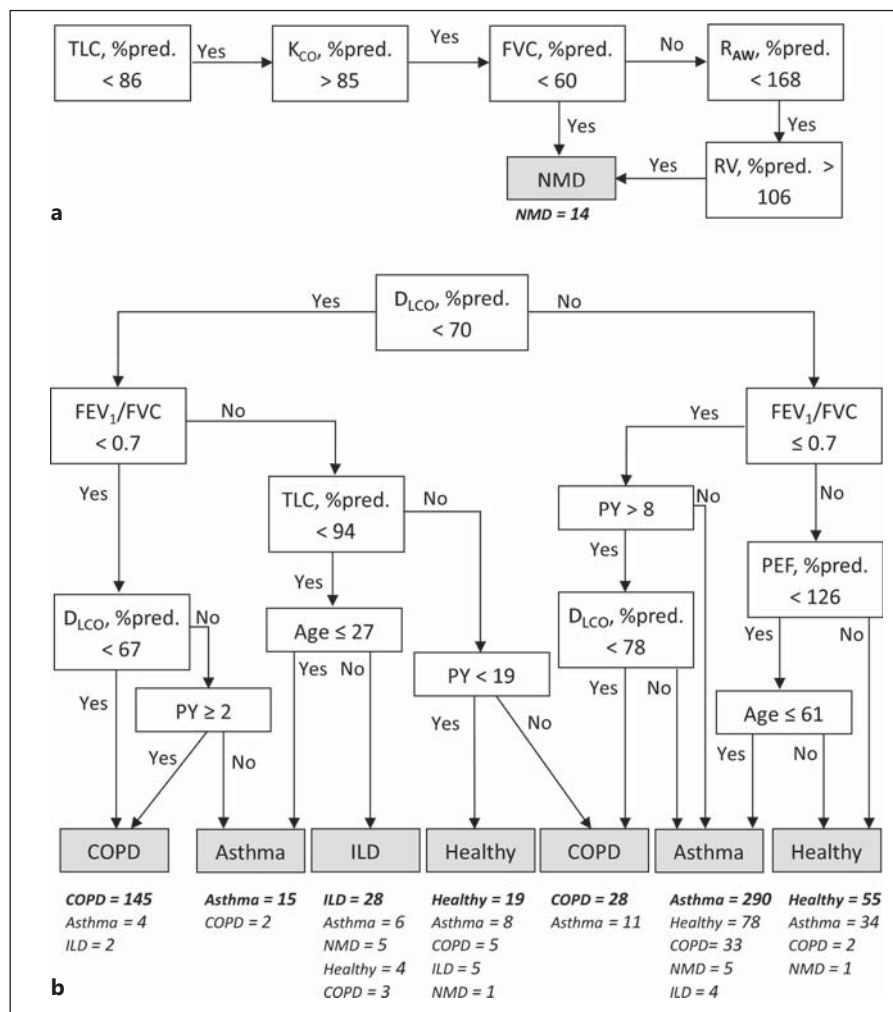
## Discussion

In this study, we developed a computer program for the automated data-driven interpretation of PFTs. When based on the international guidelines, the developed program was able to provide a correct diagnosis in only 38% of the studied population. Additionally, major mistakes were found in the discrimination of asthma from COPD and healthy individuals, and in the accurate labeling of restrictive diseases such as ILD and NMD. When developing an unbiased decision tree based on data mining programs that included not only lung function variables but also clinical patient characteristics, the accuracy increased significantly to 68%.

Automation in clinical practice is gaining pace. Apart from automated ECG interpretation, we are witnessing similar automations in detecting mammogram abnormalities, multiple sclerosis lesions on CT scans, or interpretation of laboratory tests [21–23]. Automated interpretation of PFTs has not yet become a clinical reality, due to the lack of accurate diagnostic guidelines to label lung disorders based on specific pulmonary function patterns. The weaknesses of current interpretation strate-

gies, which have been criticized by several experts in the past, are clearly demonstrated by our results [24–26]. One of the reasons for these weaknesses may be found in the choice of thresholds and parameters, where authors were motivated by simplicity rather than pure evidence. For example, in the ATS/ERS decision tree, a  $D_{LCO}$  below the LLN is used to differentiate between ILD and NMD. Both diseases may present with reduced alveolar ventilation (VA) leading to reduced gas transfer, but in contrast to ILD and interstitial involvement, the reduction in VA will be the main cause of reduced  $D_{LCO}$  in NMD.  $K_{CO}$  normalizes diffusing capacity for VA in case of chest wall disorders or NMD and is therefore a more reliable marker for the further distinction of ILD from NMD [27, 28]. When analyzing our data using the LLN of  $K_{CO}$  as a threshold, none of the NMD patients had low  $K_{CO}$ , whereas 63% of the ILD subjects did. Further analysis identified  $K_{CO}$  as the best differentiator between NMD and ILD, as a threshold at 85% of predicted values discriminated 88% of all NMD and ILD. Another weakness of the ATS/ERS decision tree lies in the definition of normal healthy lung function patterns, which is solely based on an  $FEV_1/FVC$  ratio, a VC, and a  $D_{LCO}$  above the LLN. Many lung dis-

**Fig. 5. a** A priori lung function pattern specific to discriminate subjects with neuromuscular disease (14 out of 26), yet not detecting any other disease. **b** Newly developed decision tree to assess pulmonary function and predict the 4 most common lung states. Numbers of subjects belonging to each classification group are indicated. Twelve of 26 patients with NMD are misclassified by definition. COPD, chronic obstructive pulmonary disease;  $D_{LCO}$ , diffusing capacity for carbon monoxide;  $FEV_1$ , forced expiratory volume in 1 s; FVC, forced vital capacity; ILD, interstitial lung disease;  $K_{CO}$ , transfer coefficient for carbon monoxide; NMD, neuromuscular disorder; PEF, peak expiratory flow; PY, pack-years;  $R_{AW}$ , airway resistance; RV, residual volume; TLC, total lung capacity.



eases may still appear with large disturbances of other parameters in combination with changes on these key parameters that are statistically within the normal limits, but in combination may clearly suggest disease.

By the development of a new unbiased decision tree, we are demonstrating that careful data-based modelling improves the accuracy of decision-making. Interestingly, the computer integrates clinical variables in the decision process, which is exactly what clinicians would do when reading and interpreting lung function data. For instance, the different pathways that are given by the computer are all tracks that make clinical sense, with computer-based cut-offs and thresholds that are different from the lower or upper limits that are normally used. Being data-driven, our decision tree follows clinical reasoning, but also takes into account statistical chances for a certain outcome. For example, normal  $D_{LCO}$  will never lead to a diagnosis of COPD, but rather to one of obstructive asthma. It does not

mean that COPD with normal  $D_{LCO}$  does not exist, yet the chance is low and therefore accepted as a mistake. Overall, the decision tree designed by the computer is approaching the accuracy of the expert panel, which reached 80% accuracy for healthy, asthma, ILD, NMD, and COPD based on the combination of 4 tests with clinical history [6].

Although the current algorithms are improving lung function interpretation, there are still many mistakes with asthma and its differentiation from COPD and healthy. This problem has certainly to do with the fact that asthma may appear with a complete healthy and normal lung function on the one hand, and with an irreversible obstruction mimicking COPD on the other hand. We think that further developments for automated interpretation strategies should focus on lung functions that have an a priori disturbed pattern. How this disturbance should be defined is still unclear, but it is obvious that we should be less restrictive than what the current ATS/ERS rules are

suggesting. One challenging approach may be the inclusion of new more selective parameters of PFTs to improve the differentiation between these diseases [29, 30]. Another problem for the development of automated interpretation of lung function is its inherent dependency on the prevalence of the diseases comprised in the dataset. For instance, when developing a tree with no a priori exclusion of NMD, the PPV was only 42%, with only 50% of these patients being correctly labeled despite a very specific disease pattern in a majority of these rare cases. We solved this problem by an upfront selection of the pattern on which there is no doubt and found out that this combined approach improved the overall accuracy from 64 to 68%.

An alternative approach for diagnostic labeling is to switch from easily interpretable algorithms such as decision trees to hardly or not interpretable black box algorithms such as neural networks, support vector machines, and others, which are used in most decision support systems nowadays [31, 32]. Undoubtedly, these approaches will be beneficial in terms of diagnostic correctness [33–35]. An exploratory analysis of neural networks in our dataset already increased the overall accuracy to 82% after 10-fold cross-validation [33]. However, they are dataset specific and therefore will only be applicable once we have much larger datasets. Moreover, they will be hardly accepted by clinicians if there is no logical understanding.

Finally, and inherent to any automated interpretation of PFTs, there is the need of sufficient quality of the performed tests. Different studies have shown that poorly performed tests increase the risk of misinterpretation and misdiagnosis [36, 37]. Within-test acceptability and between-test reproducibility should be achieved by adequate coaching from pulmonary function laboratory personnel. Hence, there are many guidelines assisting for optimal quality control [15, 38, 39].

To conclude, using the simple decision algorithms to determine abnormal patterns of disease and subsequently predict specific respiratory diseases is not accurate. Our data show that the current algorithms can be improved by a computer-based choice of key decision variables and their decision-making thresholds. This technology may lead to the development of computer-aided decision support systems for the accurate interpretation of PFTs.

## Acknowledgments

We thank all pulmonologists, pulmonary function technicians, patients, and hospitals participating in the study for providing data. We thank the Belgian Society of Pneumology for their financial and moral support. Finally, we thank Mr. Vasileios Exadakty-

los and Mr. Geert Celis for technical support. This work was supported by an Astra Zeneca Chair 2013–2015. W. Janssens is supported the Flemish Research Foundation (FWO).

## BPFS Investigators

R. Vanherreweghe (Algemeen Stedelijk Ziekenhuis, Aalst); R. Deman, S. Deryke, B. Ghesyens, M. Haerens, S. Maddens (AZ Groeninge, Kortrijk); W. Bultynck, W. Temmerman, L. Van Zandweghe (AZ Sint Blasius, Dendermonde); R. De Pauw, C. Depuydt, C. Haenebalcke, V. Ringoet, D. Van Renterghem (AZ Sint Jan, Brugge); A. Carlier, D. Colle, C. De Cock, J. Lamont (AZ Maria Middelaers, Gent); P. Brancalone (CH Jolimont-Lobbes); M. Vander Stappen (CHR Haute Senne); R. Peché, P. Pierard, P. Quarré, A. Van Meerhaeghe (CHU, Charleroi); D. Cataldo, J.L. Corhay, B. Duysinx, V. Heinen, M. Naldi, D. Nguyen, L. Renaud (CHU, Liège); M. Cotils, P. Duchatelet, A. Gocmen, C. Lenclud (Clinique Louis Caty, Baudour); P. Lebrun, J. Noel (Clinique Saint-Pierre, Ottignies); A. Frémault (Grand Hôpital de Charleroi); P. Bertrand, B. Bouckaert, I. Demedts, P. Demuyndt (Heilig Hart Ziekenhuis, Roeselare); E. Frans, A. Heremans, T. Lauwerier, J. Roelandts (Imelda, Bonheiden); G. Bral, I. Declercq, I. Malysse (Jan Ypermanziekenhuis, Ieper); V. Van Damme (St. Andries Ziekenhuis, Tielt); Dr. Martinot (St. Elisabeth, Namur); P. Vandenberghe (St. Maarten Duffel, Mechelen); M. Bruyneel, I. Muylle, V. Ninane (St. Pierre, Bruxelles); P. Collard, G. Liistro, B. Mwenge Gimbada, T. Pieters, C. Pilette, F. Pirson, D. Rodenstein (UCL, Bruxelles); L. Delaunois, E. Marchand, O. Vandensplas (UCL, Mont-Godinne); W. De Backer, P. Germonpre, A. Janssens, A. Vrints (UZA, Antwerpen); M. Decramer, W. Janssens, P. Van Bleyenbergh, G. Verleden, W. Wuyts (UZ Gasthuisberg, Leuven); G. Brusselle, E. Derom, G. Joos, K. Tournoy, K. Vermaelen (UZ, Gent); P. Alexander, C. Gysbrechts (Ziekenhuis Ronse); L. Bedert, Dr. Bomans, Dr. De Beukelaar, E. De Droogh, D. Galdermans, Dr. Ingelbrecht, Dr. Lefebure, H. Slabbynck, Dr. Van Mulders, Dr. Van Schaardenburg, Dr. Verbuyst (ZNA, Antwerpen); M. Daenen (ZOL Genk).

## References

- 1 Crapo RO: Pulmonary-function testing. *N Engl J Med* 1994;331:25–30.
- 2 Brusasco V, Crimi E, Crimi P, Valenti S: Comparison between intracutaneous and prick tests in detecting sensitizations to common Mediterranean allergens. *Respiration* 1980;40:317–322.
- 3 Ferrazza AM, Martolini D, Valli G, Palange P: Cardiopulmonary exercise testing in the functional and prognostic evaluation of patients with pulmonary diseases. *Respiration* 2009;77:3–17.
- 4 Singh S, Grover V, Christie L, Charles P, Kelleher P, Shah PL: A comparative study of bronchoscopic microsample probe versus bronchoalveolar lavage in patients with burns-related inhalational injury, acute lung injury and chronic stable lung disease. *Respiration* 2015;89:19–26.



- 5 Burrows B: Pulmonary terms and symbols. A report of the ACCP-STS Joint Committee on Pulmonary Nomenclature. *Chest* 1975;67: 583–593.
- 6 Decramer M, Janssens W, Derom E, Joos G, Ninane V, Deman R, Van Renterghem D, Liistro G, Bogaerts K; Belgian Pulmonary Function Study Investigators: Contribution of four common pulmonary function tests to diagnosis of patients with respiratory symptoms: a prospective cohort study. *Lancet Respir Med* 2013;1:705–713.
- 7 Sveum R, Bergstrom J, Brottman G, Hanson M, Heiman M, Johns K, Malkiewicz J, Manney S, Moyer L, Myers C, Myers N, O'Brien M, Rethwill M, Schaefer K, Uden D: Diagnosis and management of asthma. Bloomington, MN, Institute for Clinical Systems Improvement (ICSI), 2012, pp 1–86.
- 8 Vestbo J, Hurd SS, Agustí AG, Jones PW, Vogelmeier C, Anzueto A, Barnes PJ, Fabbri LM, Martinez FJ, Nishimura M, Stockley RA, Sin DD, Rodriguez-Roisin R: Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. *Am J Respir Crit Care Med* 2013;187:347–365.
- 9 Pellegrino R, Viegi G, Brusasco V, Crapo RO, Burgos F, Casaburi R, Coates A, van der Grinten CP, Gustafsson P, Hankinson J, Jensen R, Johnson DC, MacIntyre N, McKay R, Miller MR, Navajas D, Pedersen OF, Wanger J: Interpretative strategies for lung function tests. *Eur Respir J* 2005;26:948–968.
- 10 Aikins JS, Kunz JC, Shortliffe EH, Fallat RJ: PUFF: An expert system for interpretation of pulmonary function data. *Comput Biomed Res* 1983;16:199–208.
- 11 Hankinson JL: Automated pulmonary function testing: interpretation and standardization. *Ann Biomed Eng* 1981;9:633–643.
- 12 Krumpe P, Weigt G, Martinez N, Marcum R, Cumiskey JM: Computerized rapid analysis of pulmonary function test: use of a least mean squares correlation for interpretation of data. *Comput Biol Med* 1982;12:295–307.
- 13 Kligfield P, Gettes LS, Bailey JJ, Childers R, Deal BJ, Hancock EW, van Herpen G, Kors JA, Macfarlane P, Mirvis DM: Recommendations for the standardization and interpretation of the electrocardiogram: part I: the electrocardiogram and its technology a scientific statement from the American Heart Association Electrocardiography and Arrhythmias Committee, Council on Clinical Cardiology; the American College of Cardiology Foundation; and the Heart Rhythm Society endorsed by the International Society for Computerized Electrocardiology. *J Am Coll Cardiol* 2007;49:1109–1127.
- 14 Willems JL, Abreu-Lima C, Arnaud P, van Bemmel JH, Brohet C, Degani R, Denis B, Gehring J, Graham I, van Herpen G: The diagnostic performance of computer programs for the interpretation of electrocardiograms. *N Engl J Med* 1991;325:1767–1773.
- 15 Miller MR, Hankinson J, Brusasco V, Burgos F, Casaburi R, Coates A, Crapo R, Enright P, van der Grinten C, Gustafsson P: Standardisation of spirometry. *Eur Respir J* 2005;26:319–338.
- 16 Quanjer PH; European Coal and Steel Community Working Party “Standardization of Lung Function Tests”: Standardized lung function testing: report / Working Party “Standardization of Lung Function Tests”. Luxembourg, Pergamon Press, 1983.
- 17 Quanjer PH, Tammeling G, Cotes J, Pedersen O, Peslin R, Yernault J: Lung volumes and forced ventilatory flows. *Eur Respir J* 1993;6: 5–40.
- 18 Miller A, Thornton JC, Warshaw R, Anderson H, Teirstein AS, Selikoff IJ: Single breath diffusing capacity in a representative sample of the population of Michigan, a large industrial state. Predicted values, lower limits of normal, and frequencies of abnormality by smoking history. *Am Rev Respir Dis* 1983; 127:270–277.
- 19 Venables W, Ripley B: Tree-based methods; in: *Modern Applied Statistics with S-PLUS*. New York, Springer, 1999, pp 303–327.
- 20 Kohavi R: A study of cross-validation and bootstrap for accuracy estimation and model selection. *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, vol 2. San Francisco, CA, Morgan Kaufmann Publishers Inc., 1995, pp 1137–1143.
- 21 Croxatto A, Prod'homme G, Faverjon F, Rochais Y, Greub G: Laboratory automation in clinical bacteriology: what system to choose? *Clin Microbiol Infect* 2016;22:217–235.
- 22 Popescu V, Schoonheim MM, Versteeg A, Chaturvedi N, Jonker M, de Menezes RX, Garre FG, Uitdehaag BM, Barkhof F, Vrenken H: Grey matter atrophy in multiple sclerosis: clinical interpretation depends on choice of analysis method. *PLoS One* 2016;11:e0143942.
- 23 Yeh JY, Chan SW, Wu TH: Mining breast cancer classification rules from mammograms. *J Intell Syst* 2016;25:19–36.
- 24 Baur X, Degens P, Heitmann R, Hillenbach C, Marek W, Rausch V, Ulmer WT: Lung function testing: the dilemma of predicted values in relation to the individual variability. *Respiration* 1996;63:123–130.
- 25 Enright P: Flawed interpretative strategies for lung function tests harm patients. *Eur Respir J* 2006;27:1322–1323.
- 26 Kreider ME, Grippi MA: Impact of the new ATS/ERS pulmonary function test interpretation guidelines. *Respir Med* 2007;101:2336–2342.
- 27 Hughes J, Pride N: In defence of the carbon monoxide transfer coefficient  $K_{CO}$  (TL/VA). *Eur Respir J* 2001;17:168–174.
- 28 Kaminsky DA, Whitman T, Callas PW: DLCO versus DLCO/VA as predictors of pulmonary gas exchange. *Respir Med* 2007;101: 989–994.
- 29 Topalovic M, Exadaktylos V, Decramer M, Troosters T, Berckmans D, Janssens W: Modelling the dynamics of expiratory airflow to describe chronic obstructive pulmonary disease. *Med Biol Eng Comput* 2014;52:997–1006.
- 30 Topalovic M, Exadaktylos V, Peeters A, Coolen J, Dewever W, Hemeryck M, Slagmolen P, Janssens K, Berckmans D, Decramer M, Janssens W: Computer quantification of airway collapse on forced expiration to predict the presence of emphysema. *Respir Res* 2013;14: 131.
- 31 Shan J, Alam SK, Garra B, Zhang Y, Ahmed T: Computer-aided diagnosis for breast ultrasound using computerized BI-RADS features and machine learning methods. *Ultrasound Med Biol* 2016;42:980–988.
- 32 Shin H, Markey MK: A machine learning perspective on the development of clinical decision support systems utilizing mass spectra of blood samples. *J Biomed Inform* 2006;39: 227–248.
- 33 Curran SP, Mingers J: Neural networks, decision tree induction and discriminant analysis: an empirical comparison. *J Oper Res Soc* 1994;45:440–450.
- 34 Elizondo D: The linear separability problem: some testing methods. *IEEE Trans Neural Netw* 2006;17:330–344.
- 35 He H, Ma Y: *Imbalanced Learning: Foundations, Algorithms, and Applications*. New Jersey, John Wiley & Sons, 2013.
- 36 Leuppi JD, Miedinger D, Chhajed PN, Buess C, Schafroth S, Bucher HC, Tamm M: Quality of spirometry in primary care for case finding of airway obstruction in smokers. *Respiration* 2010;79:469–474.
- 37 Stoller JK, Orens DK, Hoisington E, McCarthy K: Bedside spirometry in a tertiary care hospital: the Cleveland clinic experience. *Respir Care* 2002;47:578–582.
- 38 Wanger J, Clausen JL, Coates A, Pedersen OF, Brusasco V, Burgos F, Casaburi R, Crapo R, Enright P, van der Grinten CP, Gustafsson P, Hankinson J, Jensen R, Johnson D, Macintyre N, McKay R, Miller MR, Navajas D, Pellegrino R, Viegi G: Standardisation of the measurement of lung volumes. *Eur Respir J* 2005; 26:511–522.
- 39 Macintyre N, Crapo RO, Viegi G, Johnson DC, van der Grinten CP, Brusasco V, Burgos F, Casaburi R, Coates A, Enright P, Gustafsson P, Hankinson J, Jensen R, McKay R, Miller MR, Navajas D, Pedersen OF, Pellegrino R, Wanger J: Standardisation of the single-breath determination of carbon monoxide uptake in the lung. *Eur Respir J* 2005;26:720–735.