Classification of cystic fibrosis lung disease patterns based on CT texture analysis and machine learning

Author: Pietro Quadri

Advisor: Prof. Andrea Aliverti

Co-advisor: Prof. Francesca Pennati

Academic year: 2021-2022

1. Introduction

1.1. Cystic Fibrosis

Cystic fibrosis (CF) is an autosomal recessive disease caused by the mutation of the Cystic fibrosis transmembrane conductance regulator (CFTR) protein, a ionic channel that modulates the secretion of chloride inside of the body [1]. In healthy subjects, proper mucociliary activity is promoted by the equilibrium of sodium absorption and sodium secretion. A mutation of the CFTR channel compromises the chloride secretion and lead to fluid hyper-absorption due to the osmotic pressure generated. The mucus produced in CF patients is dense and sticks to the airways surface , limiting the mucociliary clearance. Mucin plaques and plagues cause airflow obstruction and they also become site of chronic cycles of infections and inflammations leading to structural damages of the airways.
The "gold-standard" for the evaluation and monitoring of CF lung disease are pulmonary function tests (PFTs) and high resolution computed tomography (HRCT). PFTs provide a global measure of pulmonary function, while HRCT determines the severity and the spatial distribution of structural abnormalities [2].

Multiple scoring systems are present in literature for the semi-quantitative assessment of chest CT in CF lung disease [3–5] and most of them include the evaluation of bronchiectasis, airway-wall thickening, mucous plugging and parenchymal opacities. Nevertheless, the scoring of HRCT is performed by trained radiologists, thus is time consuming and introduces inter- and intra-subject variability in the evaluation of the images.

1.2. Aim

The aim of this work is to develop an automated algorithm that, based on lung CT imaging, is able to quantitatively evaluate radiological pat- terns in patients affected by CF lung disease. To this aim, healthy and pathological patterns were described through standard texture analysis features and features related to airways morphology and positioning, which represent common features used by radiologists to evaluate cystic fibrosis (CF) lung disease. The information retrieved for each pattern was used to develop a classifier, to quantify pathological patterns within the lungs in patients diagnosed with CF. Then, the correlation between the quantification of the structural damage and

pulmonary function tests was evaluated to investigate structure-function relationship in CF lung disease.

2.   Materials and Methods

2.1.   Dataset

The dataset (Table 1) was composed of 472 regions of interest (ROIs) divided, according to a radiologist, in 4 classes: 79 airway wall-thickening (AWT), 142 bronchiectasis (BR), 86 mucous plugging (MP) and 165 healthy (HLT). Two datasets with ROIs of different sizes (16x16 pixels and 32x32 pixels) were created to assess the dimension which provided the best performance.

Class	Samples
AWT	79
BR	142
MP	86
HLT	165
Total	472

Table 1: Composition of the dataset.

2.2.   Features

Each ROI was characterized with standard features based on thresholds, first order statistics and second order statistics (GLCM). In addition to these common textural descriptors, also features related to airways morphology and positioning were computed.

Standard features.   Four thresholds were selected to enhance different patterns: air-trapping, consolidation, ground-glass opacity (GGO) and high attenuation structures. The first order statistics included: mean, standard deviation, minimum, maximum, median, mode, skewness and kurtosis. Before computing the GLCM, each ROI was quantized in 16 levels, ranging from 0 to 15. Then, from the GLCM, the following features were computed: Contrast, Dissimilarity, Homogeneity, ASM, Energy, Correlation.

Airways features.   Two approaches were developed to segment the airways within a ROI.

The first one searched for contours around a manually selected HU level. An airway was accepted if it was a closed contour, which enclosed a low-density area (airway lumen) and was completely surrounded by high attenuating structures (airway wall). As this method required the constant tweaking of one parameter to correctly identify the airways, it was not adapt for an automated approach. To solve this problem, the second approach used a UNET for the identification of the airways. To create the train, validation and test-set, 10373 ROIs were extracted from CT scans of 19 patients: 17 were used for the train-set (9305 ROIs), 1 for the validation-set (490 ROIs) and 1 for the test-set (578 ROIs). Once the UNET was trained, it provided airways candidates that had to follow the same criteria of the manual method in order to be validated. Thus, for each ROI, the following parameters were extracted: the number of airways, airways average roundness, perimeter, area and the number of pixels in the prediction which had a value above 0.5. The last variable was added to account for airways partially included in the ROI.

2.3.   Features Selection

The features deemed redundant or irrelevant were removed from the dataset. Pearson correlation coefficient (ρ) was computed for each pair of features and in highly correlated features ($\rho > 0.9$) one of the two was removed. Then, the correlation coefficient was computed between each feature and the target class, retaining only features moderately correlating with the target ($\rho > 0.4$). From 24 features, only 8 were selected (Table 2).

Method	Feature
Threshold	GGO
First Order	Mean Standard deviation Skewness
GLCM	Contrast
Airways and Lung	Number of airways Average area Presence of Airways

Table 2: Final set of features.

2.4. Classification

Two classifiers, the support vector machine (SVM) and the gaussian naïve bayes (GNB), were implemented for comparison. The features dataset was divided into train (75%) and test-set (25%). As the training set was imbalanced, Synthetic Minority Over-sampling Technique (SMOTE) was applied to make the classes homogeneous.

The following metrics were used to evaluate the models' performance: specificity, sensitivity, precision and f1. The training was performed 100 times to obtain a more reliable estimate of the metrics, which are expressed as mean ± standard deviation. The process of features extraction, creation of the train and test-set and classifier evaluation is shown in Figure 1.
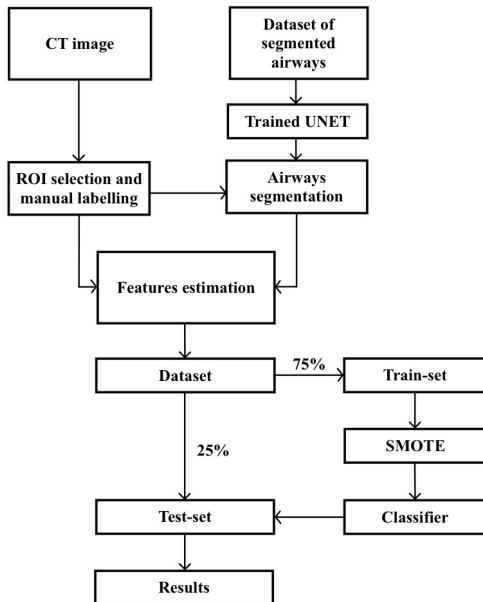

Figure 1: Dataset creation and classifier implementation.

2.5. Patient Evaluation

For each patient, the whole lung CT volume was analyzed. First, the lungs are segmented and then the vessels are removed with a Frangi's filter to limit their impact during the classification. A grid is built from the lungs bounding box and each cell presents a dimension of 8x8 pixels and is selected as a trade-off between resolution of the classification and computational cost. Then, a window slides across all the grid cells and only the ones that contained at least 80% of pixels

belonging to the lungs are used for the evaluation. Thus, each grid cell is assigned to a class. Figure 2 shows the volume evaluation process. For each HRCT, the percent volume of AWT,
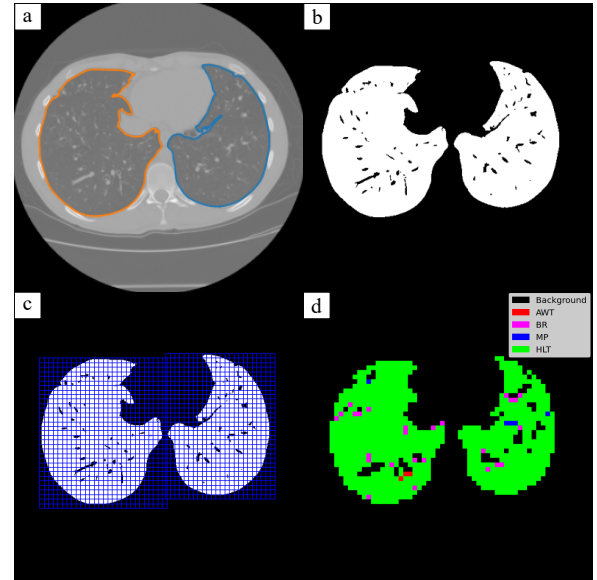

Figure 2: **(a)** HRCT image. **(b)** The lungs are segmented and the vessels are removed. **(c)** A grid is built from the bounding box of the lungs. **(d)** Each cell is assigned to one of the four classes by the classifier.

BR, MP and HLT is quantified (AWT%, BR%, MP%, HLT%). Moreover, also the percentage volume of airways is reported (vol_a%).

3. Results

3.1. UNET Evaluation

The performance of the UNET to correctly segment the airways was tested on 100 ROIs. The performance of the model was scored as follows: correct if all the airways of the ROIs were correctly segmented, non correct if at least one of the airway contained in the ROI was not correctly segmented. The UNET showed an overall accuracy of 93%.

3.2. Classifier Performance

The SVM and GNB performances were tested on ROIs of 16x16 pixels. The SVM showed a better f1 score than the GNB (SVM: 87.06 ± 2.54%, GNB: 82.32 ± 3.24%). SVM had a better performance than GNB also considering the other metrics (Figure 3). Thus, SVM was selected for the next steps.
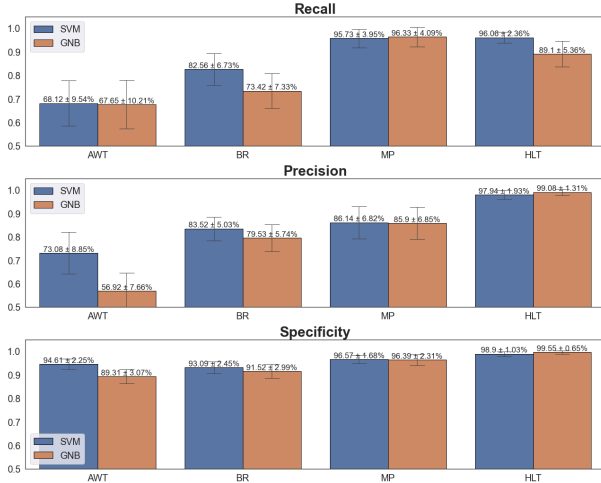
Figure 3: Comparison of SVM and GNB models

The SVM was trained on ROIs of size 16x16 pixels and 32x32 pixels to assess the impact of ROI dimension on the classification task. The model trained on smaller ROIs had a better f1 score, precision, sensitivity and specificity (f1: $87.06 \pm 2.54\%$, $83.84 \pm 2.99\%$, respectively for 16x16 and 32x32 ROIs).

To test the impact of airways features, two feature datasets were created: the first one included airways features, while the latter did not. The presence of airways features greatly improved the performance of the classifier (f1: $87.06 \pm 2.54\%$, $76.93 \pm 3.52\%$, respectively for model with and without airways features). The airways' features significantly improved the model performance for BR, AWT and surprisingly also for MP (Figure 4).
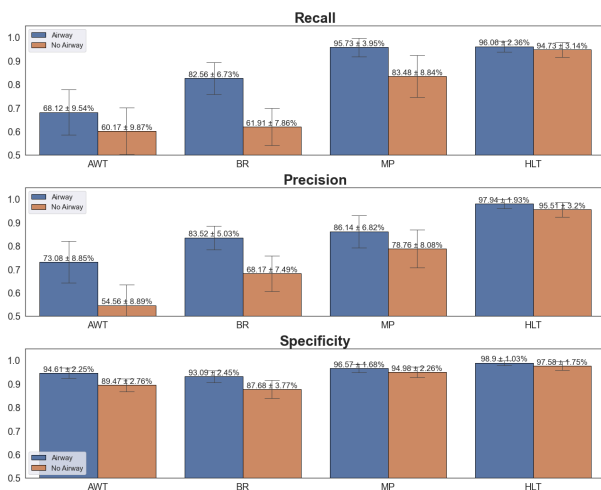

Figure 4: Comparison of the SVM trained on the dataset with and without airways features.

3.3.   Volume Analysis Results

In Figure 5 two representative patients with increasing structural damage are shown at three lung levels.

Significant correlations (Pearson) were found between the automatic algorithm quantification and the radiological scores of AWT, BR and MP (Table 3).

Scores	ρ AWT%	ρ BR%	ρ MP%
AWT	**0,72****	-	-
BR	-	**0,70****	-
MP	-	-	**0,55***

Table 3: Correlations results between radiological scores and percent volume quantified by the automatic algorithm. Significant values (*: p-value $< 0,05$; **: p-value $< 0,01$) are highlighted in bold.

Moreover, significant correlation (Spearman) was also found with some PFTs (Table 4).

4.   Discussion

Our work developed a classification model that could evaluate the whole lung volume. The analysis of an overall HRCT requires about 5 minutes.

The trained model was able to correctly identify pathological patterns, as shown from the good correlation with radiological scores.

The study population was limited, as only 15 HRCTs had radiological scores and 15 had PFTs (partially overlapped with the first group). Moreover, the patients evaluated in the present study were all enrolled in a clinical trial, thus the patients are not completely representative of the range of severity of cystic fibrosis lung disease. Indeed, only mild and moderate cases are present, as end-stage patients were not included in the trial. The algorithm should be applied to a larger cohort to have more significant results, and a bigger ROI dataset could contribute to improve the model performance.

The actual workflow analyzes 3D volumes but does not use 3D information during the classification. Texture descriptors must be adapted to work in three dimensions, and also the UNET must be modified. A 3D UNET could improve
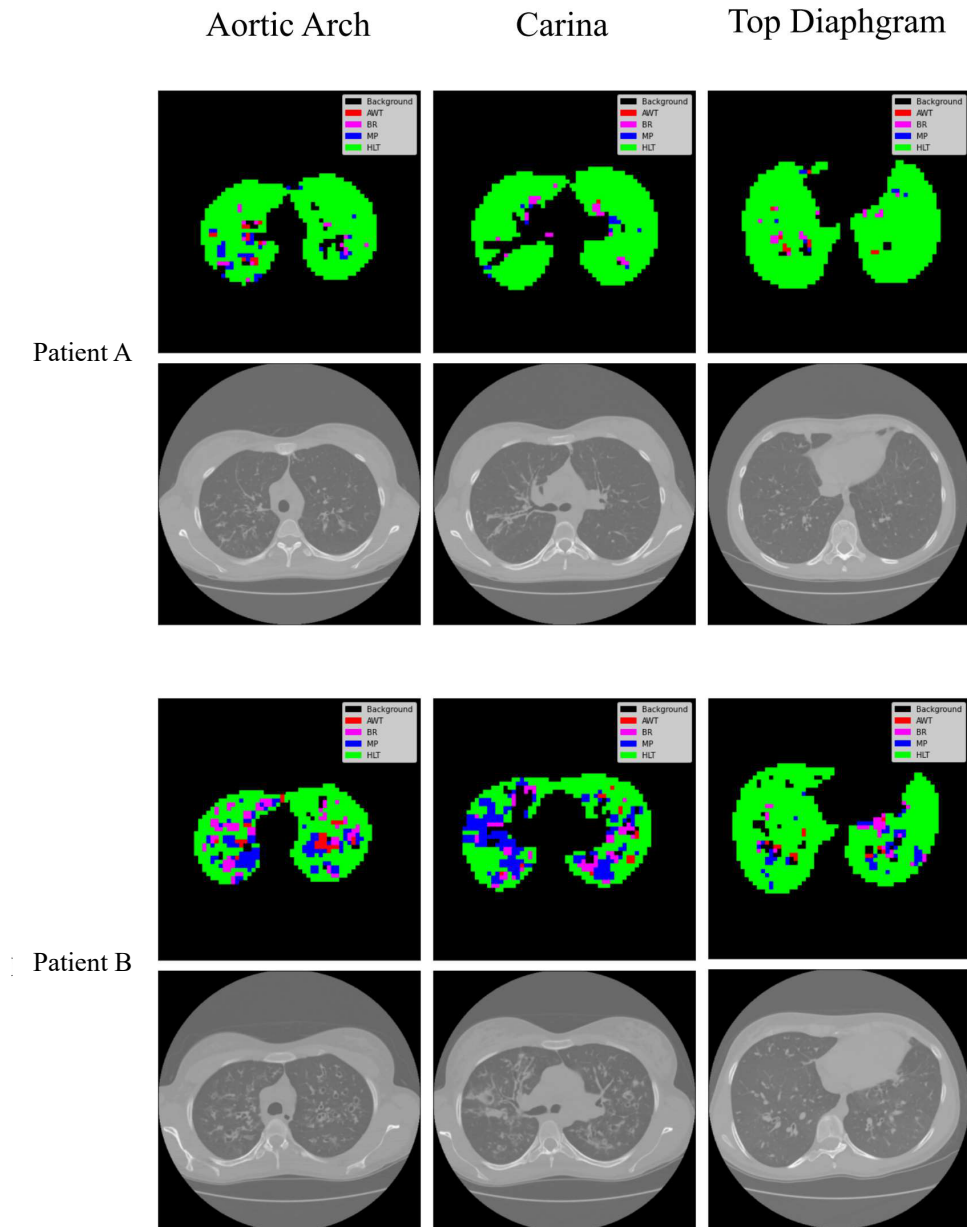
Figure 5: The images of the two patients are taken at significant positions: the aortic arch, the carina and the top diaphragm.

PFTs	ρ AWT%	ρ BR%	ρ MP%	ρ HLT%	ρ vol_a%
FEV$_1$ (L)	-0,05	-0,50	-0,48	**0,58***	-0,18
%FEV$_1$	-0,02	**-0,53***	-0,18	**0,52***	-0,47
FVC (L)	-0,03	-0,32	-0,47	0,49	-0,14
%FVC	-0,22	**-0,60***	-0,47	**0,68****	**-0,52***
FEF$_{25\%\text{-}75\%}$ (L)	-0,03	-0,36	-0,18	0,38	-0,15
%FEF$_{25\%\text{-}75\%}$	-0,02	-0,37	-0,02	0,33	-0,30
LCI 2,5% norm	0,15	0,24	0,02	-0,25	0,46

Table 4: Correlations results between PFTs and the presence of patterns in HRCT. Significant values (*: p-value < 0,05; **: p-value < 0,01) are highlighted in bold.

the segmentation of the airways, as the airways continuity could be used to differentiate them from other portions of the lungs. Moreover in a 3D approach, the UNET could be used to segment also the airway-wall and the airway complementary artery. This could be beneficial to the correct classification of AWT% and BR% as some important indexes (airway to artery ratio, airway tapering and wall to artery ratio) could be computed. These features were not included in the actual work as it was not possible to obtain a reliable measure.

5. Conclusions

This work developed an automated algorithm to evaluate HRCT in patients affected by cystic fibrosis lung disease. The algorithm provided promising results, but it need further validation in a larger cohort of patients. Further work should be aimed to improve the model performance, which might be achieved with a 3D analysis or introducing other features typical of CF.

References

[1] Nelson L Turcios. Cystic fibrosis lung disease: An overview. *Respiratory Care*, 65(2):233–251, November 2019.

[2] Alistair D. Calder, Andrew Bush, Alan S. Brody, and Catherine M. Owens. Scoring of chest CT in children with cystic fibrosis: state of the art. *Pediatric Radiology*, 44(12):1496–1506, August 2014.

[3] Alan S. Brody, Jeffrey S. Klein, Paul L. Molina, Joanne Quan, Judy A. Bean, and Robert W. Wilmott. High-resolution computed tomography in young patients with cystic fibrosis: Distribution of abnormalities and correlation with pulmonary function tests. *The Journal of Pediatrics*, 145(1):32–38, July 2004.

[4] Tim Rosenow, Merel C. J. Oudraad, Conor P. Murray, Lidija Turkovic, Wieying Kuo, Marleen de Bruijne, Sarath C. Ranganathan, Harm A. W. M. Tiddens, and Stephen M. Stick. PRAGMA-CF. a quantitative structural lung disease computed tomography outcome in young children with cystic fibrosis. *American Journal of Respiratory and Critical Care Medicine*, 191(10):1158–1165, May 2015.

[5] M Loeve, P T. W van Hal, P Robinson, P A de Jong, M H Lequin, W C Hop, T J Williams, G D Nossent, and H A Tiddens. The spectrum of structural abnormalities on CT scans from patients with CF with severe advanced lung disease. *Thorax*, 64(10):876–882, June 2009.