POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

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

Tesi di Laurea Magistrale in
Biomedical Engineering - Ingegneria Biomedica

Author: **Pietro Quadri**

Student ID: 964139
Advisor: Prof. Andrea Aliverti
Co-advisors: Prof. Francesca Pennati
Academic Year: 2021-22

Abstract

Lung computed tomography (CT), with established scoring systems, is considered the reference imaging technique in CF lung disease to detect early structural alterations, with the advantage over pulmonary function tests (PFTs) to provide information on the regional distribution of the disease. Nevertheless, scoring systems are semi-quantitative and are not routinely applied in most clinics because they are time-consuming, subjective and require training. The aim of this work is to develop an automated algorithm to detect and quantify structural abnormalities in CT scans in patients affected by cystic fibrosis lung disease. The target is to distinguish various radiological patterns typical of this pathology: bronchiectasis, airway wall-thickening and mucus plugging.

To this aim, pathological and healthy patterns were characterized through texture analysis, and a classifier was trained and tested on a dataset of manually delineated ROIs. The dataset was composed of 472 ROIs divided in 4 classes as follows: 79 airway wall thickening, 142 bronchiectasis, 86 mucous plugging, 165 healthy regions. Each ROI has been described with both standard features, i.e. first and second order statistics, and features related to the airways morphology and positioning, by implementing a supporting UNET for airways segmentation. Both a Bayesian and a Support Vector Machine (SVM) classifier have been implemented for comparison and different settings (ROIs size, features, hyper-parameters) have been tested to maximize algorithm performance. Each pattern was quantified in the whole lung in a group of patients with CF lung disease by sliding a window of 16x16 pixels and classifying the central 8x8 pixels and the results of this quantitative analysis were correlated with PFTs.

The best performance was obtained for the SVM classifier (f1 score= 87.06 $\pm$ 2.54% on the test-set). The results over the whole lung showed a significant correlation between the output of the model (BR%, AWT%, MP%) and the radiological scores. Moreover, significant correlations were found with PFTs.

Keywords: Cystic Fibrosis, Texture Analysis, UNET, Machine Learning, Computed Tomography

Abstract in lingua italiana

La tomografia assiale computerizzata (TAC) toracica, con protocolli di valutazione definiti, è considerata la tecnica di imaging di riferimento per l'individuazione di anomalie strutturali precoci in fibrosi cistica polmonare, con il vantaggio rispetto ai test di funzionalità polmonare di fornire informazioni riguardo alla distribuzione spaziale della patologia. Tuttavia, i protocolli di valutazione sono semi-quantitativi e non sono effettuati regolarmente nella maggior parte delle cliniche poiché richiedono tempo, allenamento e sono soggettivi. Lo scopo di questa tesi è d'implementare un algoritmo completamente automatizzato per identificare e quantificare anomalie strutturali in TAC di pazienti affetti da fibrosi cistica polmonare. L'obiettivo è quello di distinguere vari pattern radiologici tipici di questa patologia: bronchiectasia, inspessimento delle pareti delle vie aeree e presenza di muco.

A questo scopo, pattern patologici e fisiologici sono stati caratterizzati tramite analisi di texture, e un classificatore è stato alleato e testato su un dataset di regioni di interesse (ROIs) manualmente delineate. Il dataset è composto da 472 ROIs divise in 4 classi come segue: 79 regioni con inspessimenti delle pareti delle vie aeree, 142 con bronchiectasie, 86 con presenza di muco e 165 sane. Ogni ROI è stata descritta con caratteristiche comuni, i.e. statistiche del primo e secondo ordine, e caratteristiche relative alla posizione e morfologia delle vie aeree, con il supporto di una UNET per la loro segmentazione. Un modello Bayesiano e uno basato su SVM sono stati implementati per un confronto, e differenti configurazioni (dimensione delle ROIs, caratteristiche e iperparametri) sono state testate per massimizzare la performance dell'algoritmo. Ogni pattern è stato quantificato all'interno dell'intero polmone in un gruppo di pazienti affetti da fibrosi cistica polmonare facendo scorrere una finestra di 16x16 pixels e classificando la regione di 8x8 pixels al suo centro. I risultati dell'analisi quantitativa sono stati correlati con i test di funzionalità polmonare.

La miglior performance è stata ottenuta dal classificatore basato su SVM (f1 score= 87.06 $\pm$ 2.54% sul test-set). I risultati dell'analisi sull'intero volume polmonare hanno mostrato correlazioni significative fra l'output del modello (BR%, AWT%, MP%) e i punteggi radiologici. Inoltre, sono state individuate correlazioni significative anche con i test di funzionalità respiratoria.

Parole chiave: Fibrosi Cistica, Texture Analysis, UNET, Machine Learning, Tomografia Assiale Computerizzata

Contents

1 | Introduction

In the first chapter cystic fibrosis lung disease is overviewed. The first chapter covers the description of the disorder, the alterations occurring in the lungs and the clinical tests and imaging techniques used for the follow-up of the patients. In the end, a short description of the texture analysis techniques used in the field of CT lung imaging is provided.

1.1. Cystic Fibrosis Lung Disease

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. In general, the CFTR mutations affect the epithelial cells of an organ, and the defects in organ function are diverse, depending on their native function. In this chapter, we will focus on the lung disease in CF, which represents the major site of mortality and morbidity in this disease. For a more exhaustive description of CF pathogenesis, the reader can refer to recent reviews [1–4]. At birth, lungs of CF patients are apparently normal, before the onset of chronic, unrelenting bacterial infections of the airways in the first few years of life [5]. What is the relationship between the genetic defect and the clinical manifestations of the disease that leads to chronic respiratory infection? The CFTR ionic channel is of utmost importance to guarantee the correct mucociliary clearance of the lungs, as it acts on the equilibrium between chloride secretion and sodium absorption [4], as shown in Figure 1.1. The presence of a mutation in the CFTR protein leads to the disruption of this balance: while there is a healthy path for the sodium, the chloride secretion is compromised, causing fluid hyper-absorption due to the generated osmotic pressure. In healthy subjects, a well-balanced epithelial sodium absorption and secretion of chloride anions lead to the correct hydration of mucus and airways surface, promoting a proper mucociliary activity. In CF patients affected, the defective CFTR channel and increased activity of the epithelial sodium channel generates dehydrated mucus, which adheres to the airways surfaces and limits the mucociliary clearance. The mucin plugs and plaques that are formed with this process not only lead to air-flow obstruction, but they also become site of chronic

cycles of infections and inflammations, which result in damages of the structural integrity of the airways and lead to the development of bronchiectasis [4]. On high resolution computed tomography (HRCT) bronchiectasis, airway wall-thickening and mucus plugging are recognizable. Bronchiectasis is defined as an irreversible and abnormal dilation of the airways with reduced tapering in comparison to healthy airways and it is caused by the cycles of inflammations and infections described above [6, 7]. Airway wall-thickening is a condition in which the airway wall widens until it overcomes the 20% of the outer airway diameter or it is more than the 33% of the relative artery diameter [7]. Mucus plugging happens when the airways lumen is filled with mucus.

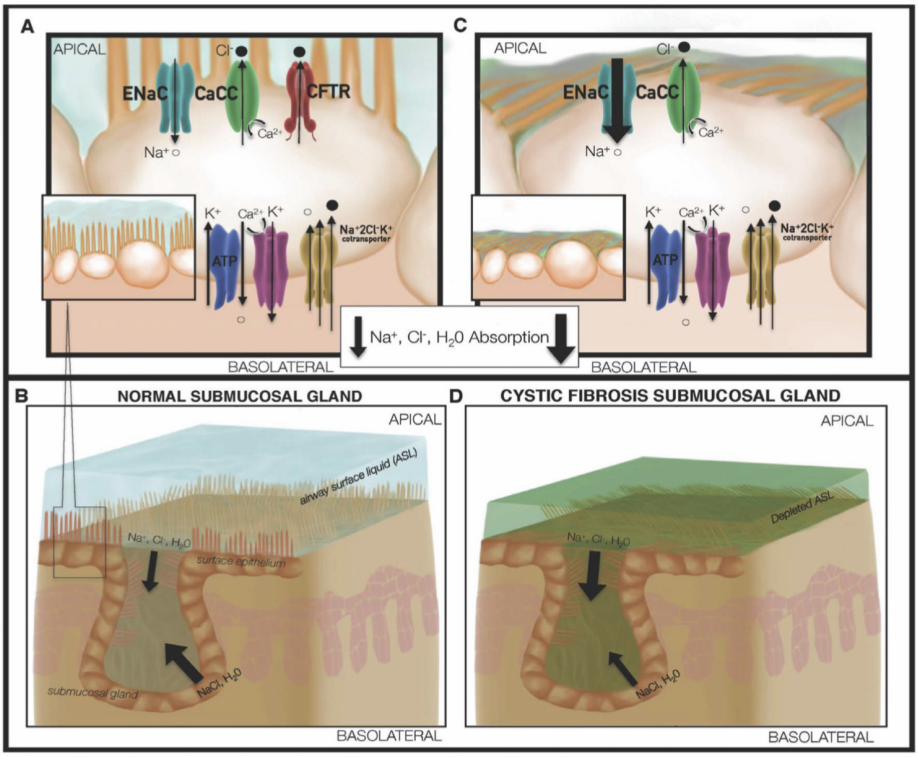


Figure 1.1: ***(a)*** The presence of normal mucous is maintained by the correct activity of CFTR, ***(b)*** therefore the airway surface is properly hydrated. ***(c)*** The lack or malfunctioning of the CFTR lead to ionic imbalance, ***(d)*** that manifest as dense mucus and reduced airway surface hydration [8].

1.2. Pulmonary Function Tests in CF Lung Disease

Pulmonary function tests (PFTs) are routinely performed in the clinical management of CF lung disease, to assess the severity of lung disease and monitor disease progression and response to therapy.

1.2.1. Static Lung Volumes Measurement

Static lung volumes are important for the evaluation of obstructive as well as restrictive ventilatory defects. The volume of gas in the lungs at a given time during the respiratory cycle are commonly referred to as "lung volumes". Figure 1.2 summarizes the standard lung volumes and capacities.

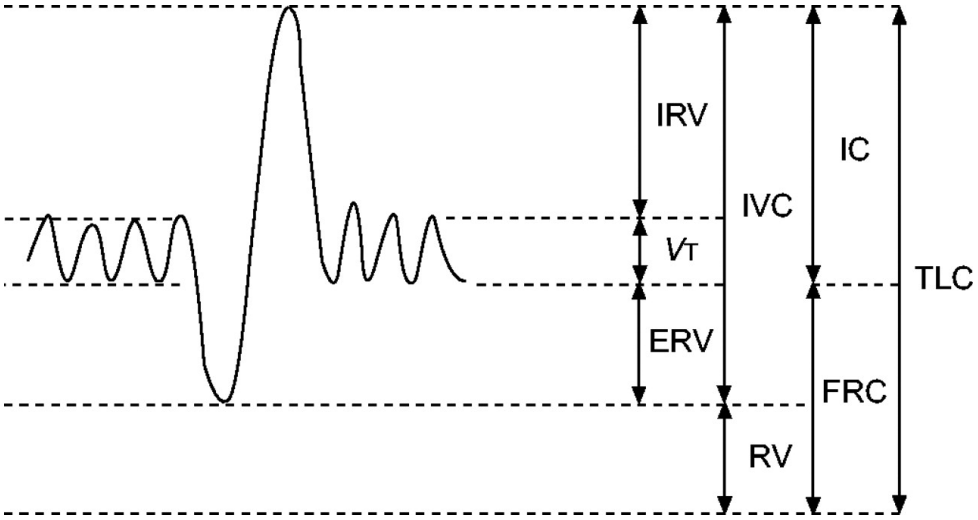


Figure 1.2: IVC: inspiratory vital capacity; FRC: functional residual capacity; ERV: expiratory reserve volume; IC: inspiratory capacity;. IRV: inspiratory reserve volume; ; RV: residual volume; ; TV: tidal volume; TLC: total lung capacity [9].

Functional residual capacity (FRC) is the amount of gas which is present inside of the lungs at the end of the normal expiration. The volume that can be forcefully inhaled from this condition is called inspiratory capacity (IC), while the one that can be exhaled is called expiratory reserve volume (ERV). After a forced expiration, the volume of gas remaining in the lung is known as reserve volume (RV). At the end of normal inspiration, the volume of gas that can still be inhaled is called inspiratory reserve volume (IRV). The volume of gas that is normally exchanged during the respiratory cycle is known as tidal volume (TV). Inspiratory vital capacity is the maximum volume of gas that can be inhaled after a complete exhalation.

The inhaled and exhaled lung volumes can be used to detect and evaluate the progression of disease. The measurement of the volumes can be done with various techniques: body plethysmography, nitrogen washout, gas dilution, radiographic imaging methods and spirometry [9].

1.2.2. Spirometry

Spirometry is the most common pulmonary function test performed in the clinical setting. It measures the rate of change of lung volumes during forced breathing maneuvers. To perform the test, a maximum inspiration followed by a forced exhalation is performed through a spirometer. The spirometer measures the airflow generated by the patient and derive the lung volume by integration. The outcomes are:

- Forced Vital Capacity (FVC): is the amount of gas that the patient is able to blow starting in a condition of full inspiration.
- Forced Expiratory Volume (FEV): is the volume of air exhaled by the patient in a defined period of time (e.g.: 1, 2, 3, ... seconds), and generally FEV_1 is used.
- FEV_1/FVC ratio.
- Forced Expiratory Flow (FEF): are the flow measured when FVC is equal to 25%, 50% and 75%.

The Figure 1.3 and Figure 1.4 show examples of data that can be obtained from of a spirometric test.

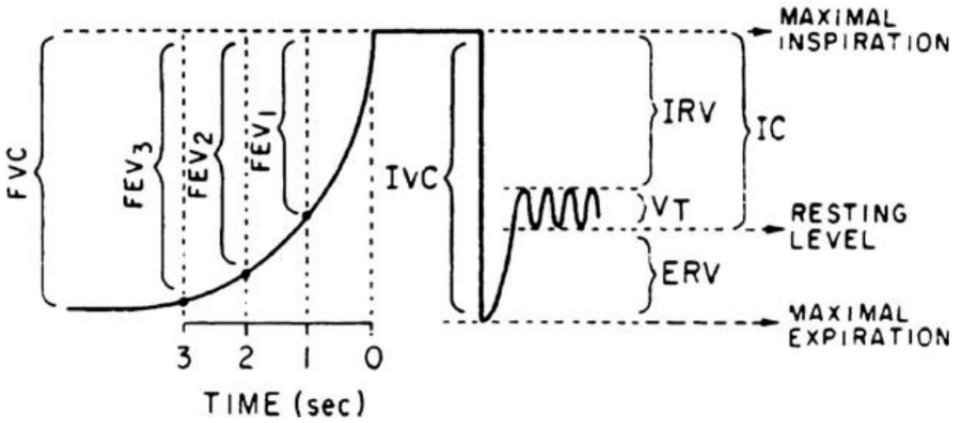


Figure 1.3: Spirometric measurement. Please note that y-axis (exhaled volume) goes from top to bottom and x-axis goes from right to left [10].

The outcomes of PFTs are strictly related to the subject which is performing the tests as lung function is affected by his age, sex, ethnicity, etc... [11]. Thus, spirometric results are compared with predicted values which are used to evaluate the health of the lungs. The predicted values are computed with the help of reference equations. These models are built on the assumption that PFTs outcomes of healthy and non-smoking subjects can be used to distinguish normal and pathological condition [11]. The normal range is

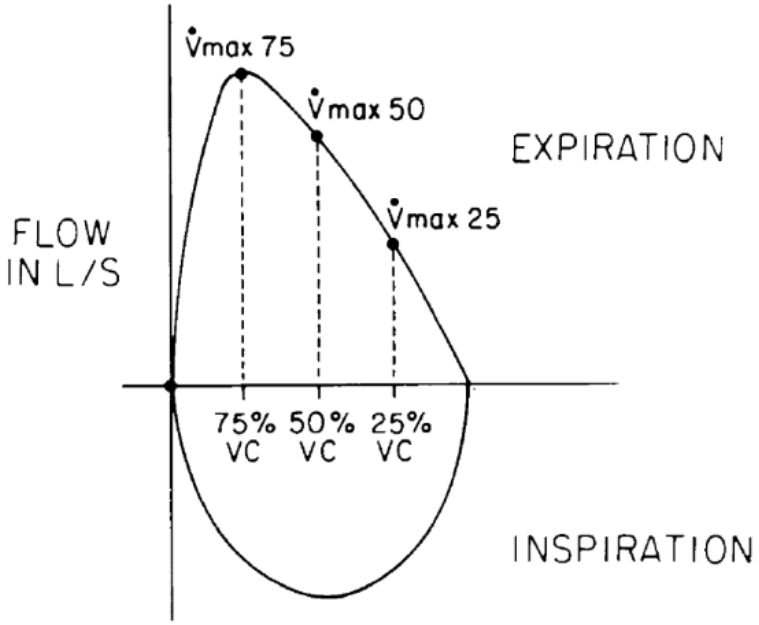


Figure 1.4: Flow-Volume curve. Please note x-axis (lung volume) goes from right to left [10].

built around the predicted value because it must reflect the variability within the healthy population. A value of 10% is often accepted as "normal" variation coefficient, therefore values in the range 80% - 120% of the predicted values are considered normal [11].

In CF lung disease, the decline rate of FEV_1 has been considered one of the most important outcomes to monitor and evaluate the progression of lung disease [12].

FEV_1 is considered in the normal range if it is more than the 80% of the predicted value. Lung disease is commonly defined mild (>60% or 70% of predicted), moderate, and severe (<40% or 30%). Moreover, FEV_1 can be used to evaluate the severity of obstructive lung disease, and in combination with FVC also restrictive lung disease. In obstructive lung disease, FEV_1 is more sensitive to changes of airway resistance in the expiratory phase than FVC. Thus, a reduced FEV_1 in presence of almost normal FVC leads to FEV_1/FVC ratio lower than 70%. Instead, in restrictive lung disease FVC is more affected than FEV_1, and this leads to a FEV_1/FVC ratio higher than 70%. FEV_1 is expected to rise during the growth and reaches a maximum around 20-25 years, and a plateau during the adulthood. Then, as a person ages, FEV_1 starts to decline which accelerates with age [13]. The FEV_1 decline process is shown in Figure 1.5.

FEV_1 is not easily measured in young patients because they can't easily sustain the effort which spirometry requires [15, 16], and it is not sensible to lung abnormalities in the early

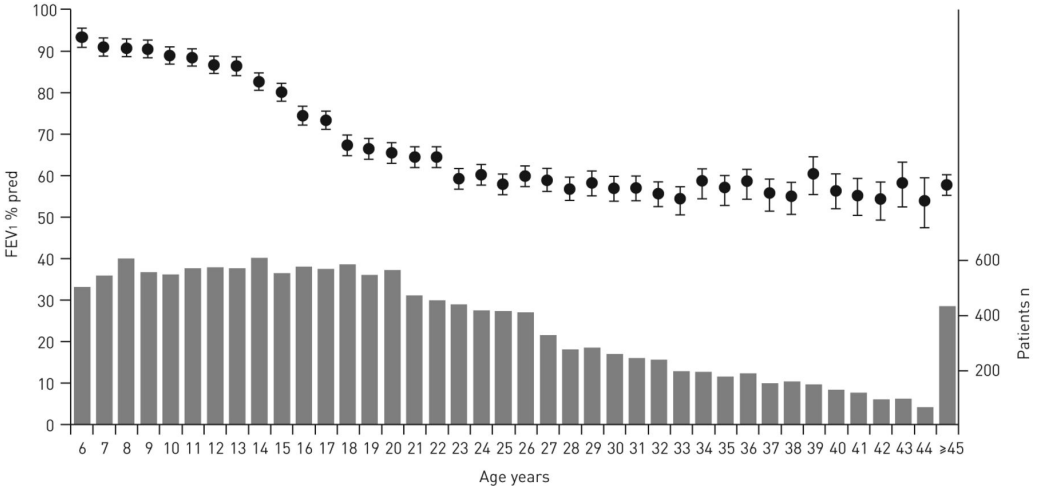
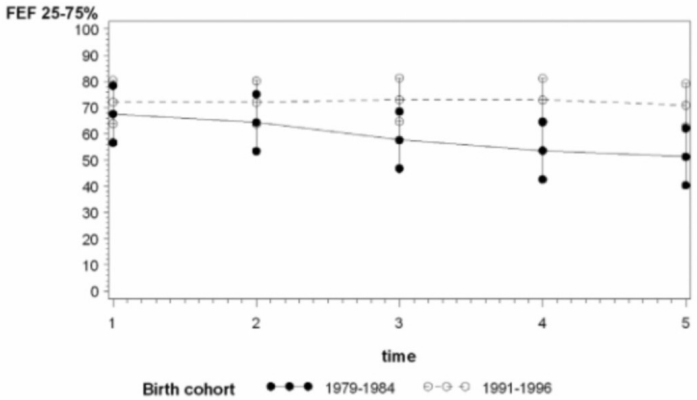


Figure 1.5: %FEV_1 predicted decline in patients affected by cystic fibrosis (95% CI). The graph shows also the number of patients by age [14].

stages of disease [4].

In recent years $FEF_{25\%-75\%}$ has gained attention as a marker of early lung disease, as more sensitive than FEV_1 to small airways obstruction [17]. $FEF_{25\%-75\%}$ is considered an indicator of small airways obstruction [18] as it is abnormal in the early stage of cystic fibrosis lung disease [19]. In Figure 1.6 is shown the comparison of $FEF_{25\%-75\%}$ decline in two cohorts of patients affected by cystic fibrosis from 10 to 15 years of age.



Birth cohort	1	2	3	4	5	Overall
1979-1984	67.6 (56.7, 78.5)	64.4 (53.5, 75.3)	57.7 (46.8, 68.6)	53.6 (42.7, 64.5)	51.3 (40.4, 62.2)	58.9 (48.7, 69.1)
1991-1996	72.2 (64.0, 80.5)	72.1 (63.8, 80.3)	73.1 (64.9, 81.3)	73.0 (64.8, 81.3)	70.9 (62.7, 79.2)	72.3 (64.5, 80.0)

Figure 1.6: $FEF_{25\%-75\%}$ decline in two cohorts of patients from 10 to 15 years of age [20].

Spirometry is an easy clinical test to monitor cystic fibrosis lung disease. However, spirometric results can remain within the normal range even in presence of anomalies in lung structure [21] and they provide a global value of lung function, without giving any information about the distribution of the abnormalities inside of the lungs.

1.2.3. Multiple Breath Washout Test

Another technique that has shown promising results in the early detection of lung disease is the multiple breath washout test (MBW). The test is composed of two phases: the wash-in phase, and the wash-out phase. In the first part of the test, the subject is required to inhale a test gas so that the air inside of his lungs reach a defined composition. The test gas can be either extrinsic, if it is not usually present inside of the lungs (SF_6 and He), or intrinsic (N_2). The wash-in stops once the gas concentration in the exhaled air is the same at which the gas is delivered. Then, during the wash-out phase, another gas is provided to the subject and it is used to "wash-out" the test gas from the lungs until it reaches 1/40 of the initial concentration as shown in Figure 1.7.

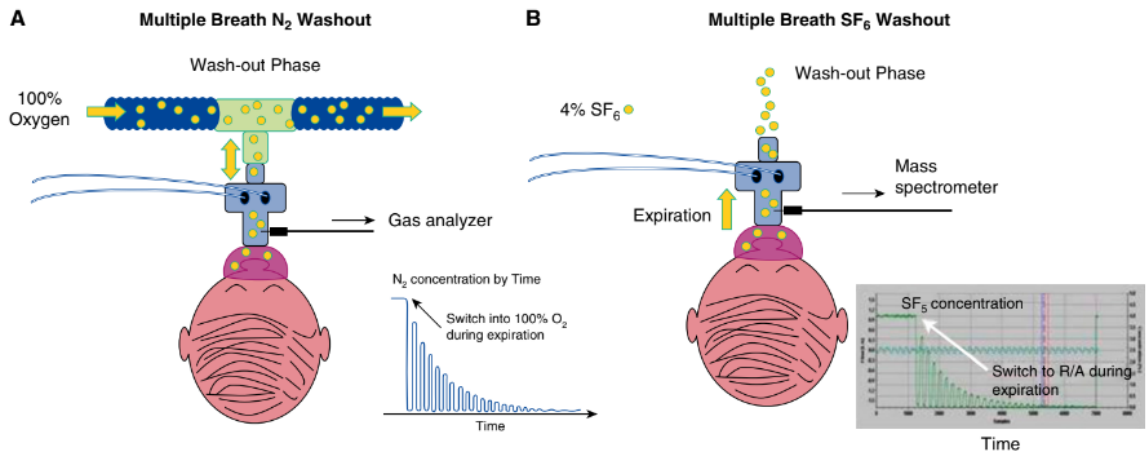


Figure 1.7: Example of wash-out phase with an intrinsic ***(a)*** and extrinsic ***(b)*** gas [21].

At the end of the test the lung clearance index (LCI) can be calculated as:

$$LCI = \frac{V_{ce}}{FRC}$$

where V_{ce} is the volume expired in order to reach 1/40 of the test gas concentration and FRC is the functional residual capacity [21].

LCI is a measure of ventilation inhomogeneity as it represents the relative ventilation needed to free the lungs from a tracer gas [22] and it increases in presence of a patho-

logical condition. As the washout phase is slower in the presence of mucus retention or inflammation, in CF lung disease LCI increases with disease severity, thus it helps to assess changes in small airways that are not visible while using spirometry [23]. In Figure 1.8, a comparison between the LCI measured in patients affected by cystic fibrosis lung disease and healthy subjects is shown.

The LCI upper limit of normal range for young children is around 7.9, and in the study of Kinghorn et al. [23], the LCI of children affected by cystic fibrosis was 24% higher than the healthy control group. The LCI has proven to be more sensitive than FEV_1 in the detection of early signs of the disease, but further evidence is required [15].

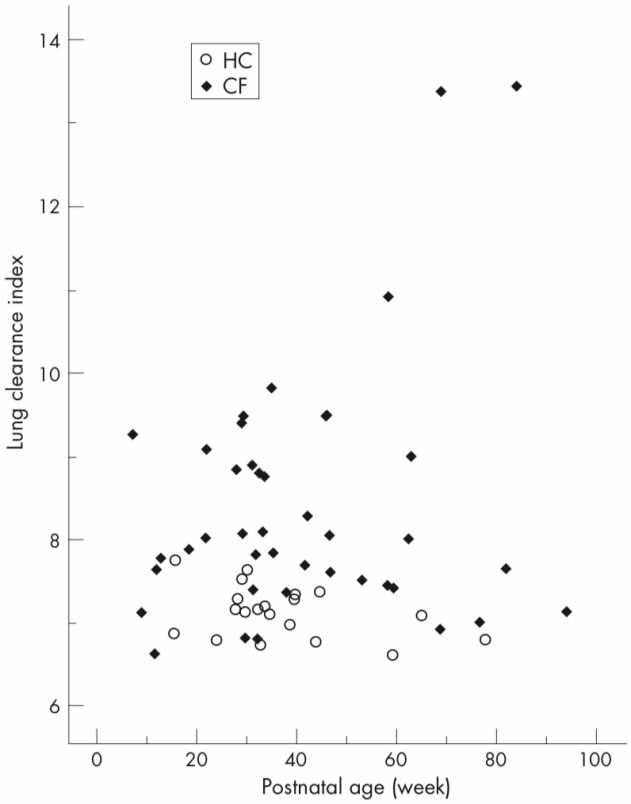


Figure 1.8: The LCI for patients affected by cystic fibrosis (CF) and control group HC is computed against the post-natal age [22].

LCI has proven to be more sensitive than FEV_1 in the early stages of the disease with results comparable with HRCT [15], but without exposing the patients to ionizing radiation. LCI is a measure of ventilation inhomogeneity, and as spirometry, it does not provide information about the structures' abnormalities in the lungs and their distribution [15].

1.3. HRCT in Cystic Fibrosis Patients

Pulmonary function testing is a useful tool to provide a fast evaluation lung function and monitor disease progression but it presents some limits: early signs of the disease might not be detected and it doesn't provide spatial information about the distribution of the disease inside of the lung [16]. Out of the different imaging techniques, High Resolution Computed Tomography (HRCT) has emerged because of its higher resolution compared to Chest X-Ray and MRI. In CF lung disease, the chronic bacterial infection and the exaggerated immune response are responsible for structural damage in the airways and parenchyma. These changes are extremely heterogeneous in the lung, with even patients with normal lung function showing lung areas with end-stage disease. HRCT is currently the "gold standard" to delineate these structural abnormalities [24–26]. It has been demonstrated that HRCT is sensitive to early disease, suggesting that structural changes occur before functional decline can be appreciated [16, 27, 28]. The most relevant structural changes occurring in CF lungs that can be observed on CT are bronchiectasis, airway wall thickening and mucous plugging. Less common structural abnormalities are atelectasis/consolidation, alveolar consolidation, sacculations/abcesse, bullae, and thickening of septa [29].

Bronchiectasis

Bronchiectasis is a permanent condition in which airways are widened as a result of damages of the surrounding lung tissue [30]. In patients affected by cystic fibrosis, the chronic cycles of inflammation and infection are responsible for the development of bronchiectasis [31], but it can also derive from airway obstruction. Bronchiectasis can be observed on HRCT as an airway with widened lumen and irregular wall, or as an airway which lacks the normal tapering, or the appearance of airways in the most peripheral areas of the lungs [31]. Airway dilation occurs when the internal diameter of the airway is larger than the diameter of the corresponding pulmonary artery, whereas airways tapering is the airway caliber decrease from the central to the peripheral areas [31]. Even if the damages are permanent and cannot be recovered, a therapeutic approach can be used to limit the progress of bronchiectasis [31]. Figure 1.9 and Figure 1.10 show the appearance of bronchiectasis in HRCT.

Figure 1.9: Severe bronchiectasis can be observed [32].

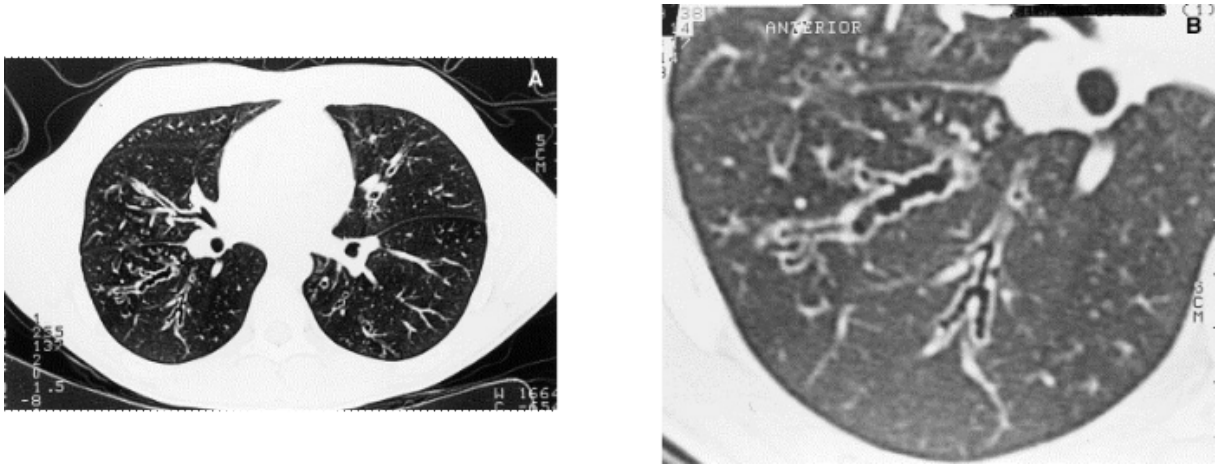


Figure 1.10: HRCT showing presence of bronchiectasis and peribronchial wall-thickening [33].

Airway wall-thickening

Airway wall-thickening, or bronchial wall-thickening, describes airways with abnormal wall thickness. This origin of this sign is thought to be the result of a process of airway modeling and inflammatory changes [34]. The thickness of the airway wall is considered pathological when it is 20% higher than the outer airway diameter or it is 33% larger than the corresponding artery diameter [7].

Airway wall-thickening can be observed in Figure 1.11.

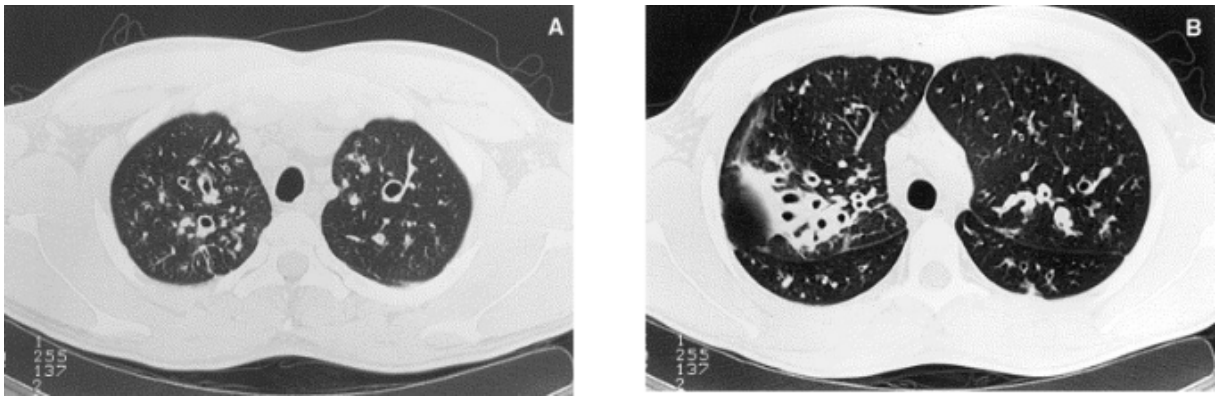


Figure 1.11: Patient affected by cystic fibrosis. Presence of bronchiectasis, airway wall-thickening and moderate mucous plugging can be observed [33].

Mucous Plugging

Mucous plugging, or mucoid impaction, is a common CT pathological finding in which airways are filled with mucous up to the point to be completely obstructed. The presence of occluded small airways can be associated with worse lung function [35]. Moreover, if the plugged airways are small enough, it could lead to collapsed alveoli and air-trapping. The mucous plugs inside of the lung are not only a problem from a mechanical point of view, as it obstructs the airways, but they can also harbor pathogens that can cause further damages to the airways and lung tissue.

Figure 1.12 show the appearance of mucous plugging in a CT scan.

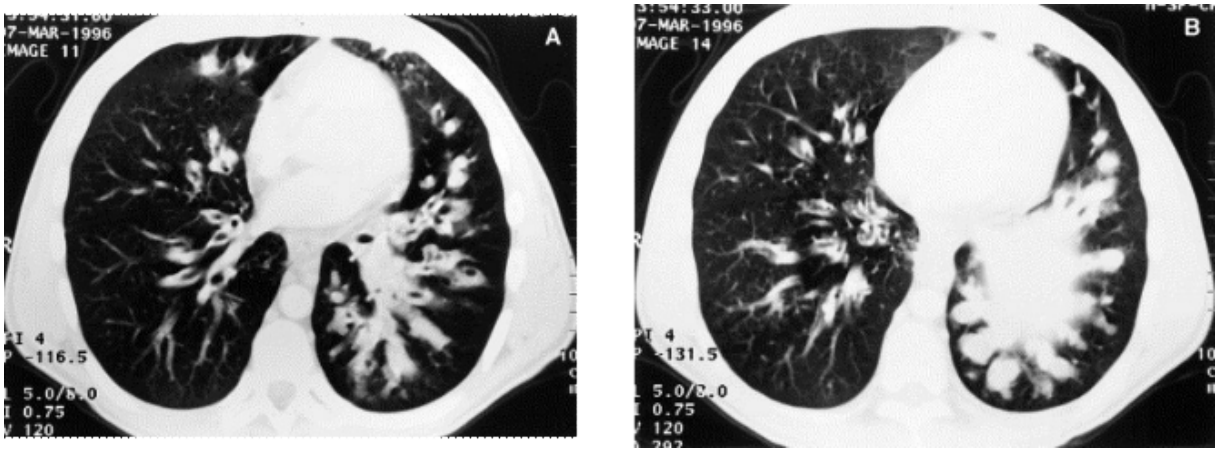


Figure 1.12: CT shows the presence of acute bronchiectasis and modest airway wall-thickening. In the left lung, extensive mucoid impaction can be seen [33].

1.3.1. Scoring Systems

Over the last few decades it has been shown that bronchiectasis, airway wall thickening, mucous impaction, and trapped air count as the most important markers to quantify on chest CTs. Numerous scoring system have been proposed for chest CT in CF lung disease (e.g.: Brody [36], PRAGMA-CF [37], SALD [38], ...) for semi-quantification of morphological CT changes. For example, bronchiectasis can be evaluated by comparison between the airway lumen and relative artery diameter (airway to artery ratio) or by checking the presence of the natural tapering of the airways and the extent is defined by the number of lungs lobes which are involved. In Figure 1.13 is shown an example image scored with PRAGMA-CF.

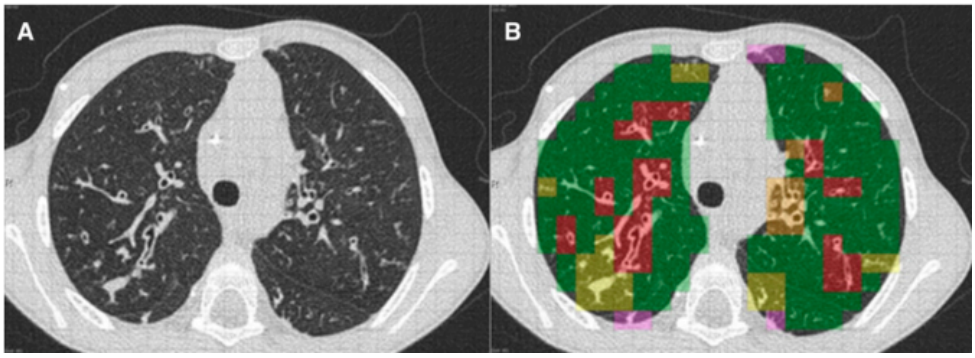


Figure 1.13: ***(a)*** The original image, with the overlaid scoring greed. ***(b)*** Slice with annotated scores: red = bronchiectasis; yellow = mucous plugging; orange = bronchial wall thickening; pink = atelectasis; green = no abnormality seen [37].

Each scoring system defines its own set of features, with bronchiectasis, mucous plugging, airway wall thickening and parenchymal opacities representing the abnormalities included in most scoring systems. Scoring systems are advantageous as they are relatively insensitive to the CT scanner technique and acquisition protocol being used.

CT scoring systems can be used to monitor patterns of disease and to identify possible risk factors in the disease progression. Bouma et al., for example, used PRAGMA-CF to show how the percentage of disease in preschool children is associated with bronchiectasis in school age children [39]. Similarly, Sanders et al., proved that CT scans scored with BRODY scoring system predict the presence of pulmonary exacerbations in the following 10 years [40]. The utility of this approach is not limited to the evaluation of the disease but could provide also a useful tool for planning lung transplantation and improving survival prediction models [41].

Even if the total score obtained with various scoring systems have shown to have a certain

degree of reproducibility, it is not clear how meaningful it is to sum scores which evaluates different patterns without knowing the weight that each one of them has on the progression of the disease [24]. In addition, this type of approach requires the training of a radiologist in the use of a specific scoring system, introducing complications which are typical of manual approaches, such as inter and intra-subject variability in the evaluation process [24]. For all these reasons, an automated approach which can identify the presence of pathological pattern based on objective measures could be an interesting tool for the support for clinical care and for the monitoring of the disease progression.

1.4. Pattern Recognition in CT Lung Imaging

Pattern recognition is a technique that exploits machine learning algorithms in order to identify patterns inside the data. In Figure 1.14 is shown the working principle of a statistical pattern recognition approach. During the training, the model selects the best set of features that classify the training samples. Then, in the test phase, the model classifies an input pattern into a class based on its features.

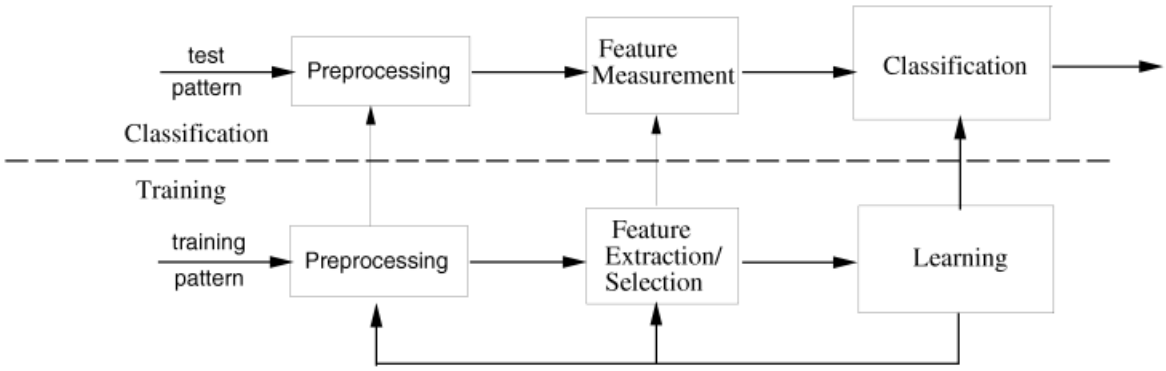


Figure 1.14: Model for statistical pattern recognition [42].

Pattern recognition applied to CT lung images can be used to extract quantitative information to monitor disease extent and progression, providing support in the decision or diagnosis process.

1.4.1. Texture Analysis

The first step in machine learning is to extract the features that contain the information used to make decisions. The simplest features derive from the application of threshold-based methods, which enhance high or low attenuation areas of the lungs [43–45]. Another common set of features are first-order statistics, which collect information that can be

derived from the histogram of the image (e.g.: mean, standard deviation, mode, kurtosis, skewness, etc...) as shown in Figure 1.15.

To exploit the spatial distribution of the data the second-order statistics, such as gray-level co-occurrence matrix (GLCM) [46] or the gray-level run length matrix are computed. An example of GLCM is shown in Figure 1.16.

Other methods rely on models or on transforms to extract more complex patterns. This class includes approaches such as Minkowski functionals, wavelets, Gabor filters and Laplacian transforms [47, 48]. This approach can be extended also for 3D regions, but their analysis falls beyond the scope of this work [49].

In Figure 1.17 the most common features used in texture analysis are summarized [48].

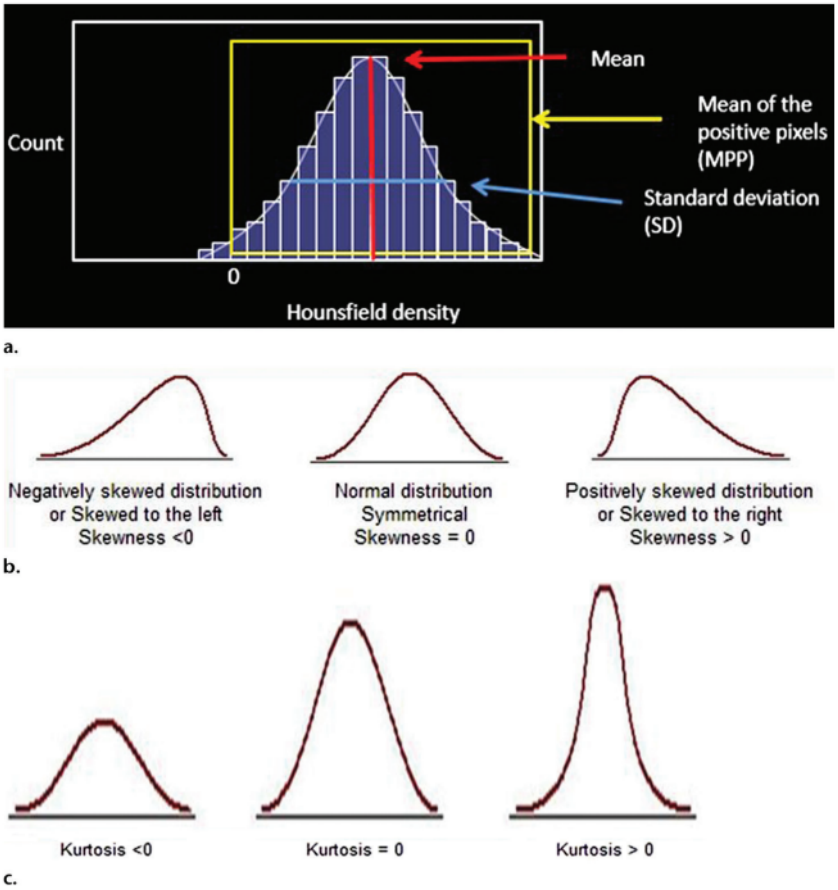


Figure 1.15: First order statistics based on a CT image. ***(a)*** Plot of the image histogram on which are underlined mean, standard deviation (SD), count of pixels and mean of positive pixels (MPP). ***(b)*** Sample histograms show the behaviour of skewness, which is a measure of asymmetry of the histogram. ***(c)*** Sample histograms show the behaviour of kurtosis, which is a measure of tailedness of the histogram. [50].

Buch et al. [52], studied the influence of the different CT acquisition parameters on textural features. In detail, they extracted 42 different features by varying pitch, milliampere, kilovolt, section thickness and acquisition mode. They concluded that even if some acquisition parameters didn't have a strong impact on the textural features, it is necessary a standardization to prevent confusion between biological features and image heterogeneity.

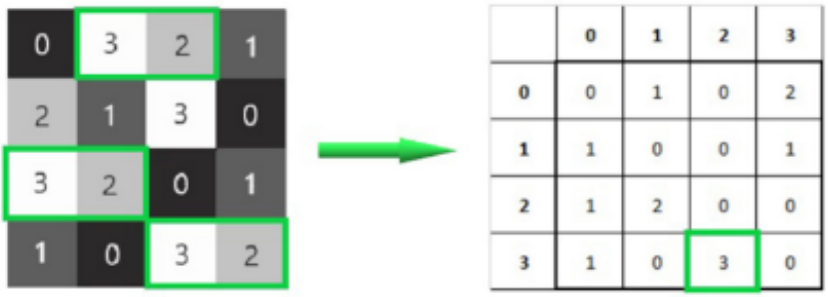


Figure 1.16: GLCM computes the occurrences of pair of pixels at a defined distance and angle in an image [51].

Class	Type	Method	Texture interpretation	Main estimated features
Statistical method	First order	Grey-level histogram	Global distribution of intensity values in terms of spread, symmetry, flatness, uniformity and randomness	Mean
				Variance
				Skewness
				Kurtosis
				Energy
				Entropy
	Second order	GLCM	Spatial relationship between pixel in a specific direction, highlighting the properties of uniformity, homogeneity, randomness and linear dependency of the image	Energy
				Homogeneity
				Contrast
				Entropy
				Dissimilarity
				Correlation
	Higher order	NGTDM	Spatial relationship among three or more pixels, closely approaching the human perception of the image	Complexity
				Busyness
				Contrast
				Coarseness
				Texture strength
	Higher order	GLRLM	Texture in a specific direction, where fine texture has more short runs whilst coarse texture presents more long runs with different intensity values	Short-run emphasis
				Long-run emphasis
				Grey-level non-uniformity
				Run length non-uniformity
				Run percentage
	Higher order	GLSZM	Regional intensity variations or the distribution of homogeneous regions	Zone size emphasis
				Grey-level zone emphasis
				Zone size non-uniformity
				Grey-level non-uniformity
Model-based method	Fractal models	Box counting, fractional Brownian motion method and power spectral method	Complexity of the image. For 2D images, FD ranges from 2 to 3, where more complex pattern presents higher values	FD
				Fractal abundance
				Lacunarity
Transform-based method	Fourier transform	Analysis of the frequency content without spatial localization		
	Wavelet and Gabor filters	Frequency and spatial localization		
	LoG	Extraction of areas with increasingly coarse texture patterns		

Figure 1.17: Features used in texture analysis, adapted from [48].

1.4.2. Texture Analysis applied to CT Lung Imaging

Texture analysis can retrieve information from CT scans that can be used as features during pattern recognition. In CT lung imaging, the pattern recognition is used to classify between pathological and healthy regions, by first extracting a set of features for each pattern, to be used to train a classification model. The results can be used to monitor and evaluate the progression of the disease.

Yoon et al [53], tried to combine the use of texture analysis with machine learning in order to classify radiological patterns (normal, alveolar, bronchial, and unstructured interstitial) in animal radiographs. They extracted 44 features, which were used to train and test an artificial neural network, that reached an area under the receiver operating characteristic of 0.92 on the test set.

Sun et al. [54], used texture analysis in order to differentiate between bronchiolar adenoma (BA), adenocarcinoma in situ (AIS) and minimally invasive adenocarcinoma (MIA). A total of 396 textural features were extracted and their study suggested that texture analysis has great potential in differentiating BA from AIS and MIA.

Tanizawa et al. [55], analyzed images of patients affected by fibrotic idiopathic interstitial pneumonias. To first order statistics parameters, they added two scores: high-attenuation areas (HAAs)(CT values $>$ -200) and low-attenuation areas (LAAs)(CT values $<$ -960) and from this cystic areas (CA) were computed. The first index reflected physiological impairment, while CA% corresponded with diagnosis and CT patterns.

Chen et al. [56], tried to use radiomic features to distinguish between benign and malignant lung nodules. They extracted for each case 750 features, but only 76 had significant differences in the two classes and the top 4 features were used to build a radiomic signature. The signature was then used to differentiate between benign and malignant lesions achieving an accuracy of 84%.

These studies suggest that a combined use of texture analysis and pattern recognition can differentiate among pathological and healthy patterns independently from the disease that generated them.

1.4.3. CT Texture Analysis in Cystic Fibrosis Lung Disease

Texture analysis and pattern recognition are not widely diffused in the evaluation of HRCT in cystic fibrosis lung disease. In the literature, the analysis of a CT image is usually limited to the extraction of some features or scores, which are then compared

with FEV_1 or scores made manually from radiologists. Nevertheless, scoring systems are not routinely applied in most clinics because they are time-consuming, subjective and require training.

Chassagnon e al. in 2018 [57], investigated the ability of a model, based on fixed and moving thresholds, to evaluate the presence of high attenuating structures inside of the lungs. The score was defined as the ratio between the lung volume above the threshold (high-attenuating structures) and the whole lung volume. A total of fourteen different thresholds were tested, four fixed and ten variable, and FEV_1 was used as gold standard. The automated score had a better correlation with FEV_1 than visual scores made by radiologist.

Instead, DeBoer et al. [58], built an automated software to count the number of visible airways and evaluate air-trapping in CT images. The scores obtained were then validated with Brody CT scan scores and pulmonary function measures. The detection of the airways was done on inspiratory images, while estimation of air-trapping on expiratory images. The number of total, central and peripheral airways in children affected by cystic fibrosis was higher than the healthy control group. Also the low attenuation density (LAD%) index, which was defined as the percentage of lung volume affected by air-trapping, was higher in children with cystic fibrosis. The number of airways correlated with Brody's bronchiectasis subscore, and LAD% with air-trapping. Hence the number of airways and LAD% could be used as surrogates in the evaluation of bronchiectasis and air-trapping in CT images.

Chassagnon et al. in 2020 [59], designed a model based on 38 radiomic features in order to assess disease severity and the risk of exacerbation in adults affected by cystic fibrosis. Out of the 38 features, 7 belonged to first order characteristics, 24 to three-dimensional co-occurence matrices and the last 7 to fractal dimensions. The extracted features were used to train a regression model that produced a radiomic based CT score. The output of the trained model showed good correlation with FEV_1 and the presence of an exacerbation in the following 12 months.

Santos et al. [60], proposed and automated method to quantitatively evaluate morphological changes of airways. In patients with severe airflow obstruction, an higher number of bronchi was found. While for patients without severe airflow obstruction the relative wall thickness of the from third bronchial generation to the eighth had a good correlation with FEV_1.

1.5. Aims

Textural analysis has proven to be useful to evaluate different lung patterns in CT images. The scope of this thesis is to create an automated algorithm that based on textural analysis and machine learning quantifies the extent of pathological patterns in patients affected by cystic fibrosis lung disease. To improve the performance of textural features, surrogates of indices used by radiologist for the evaluation of different patterns will be used. At the end, the model will be evaluated in a group of CF patients, by comparing the results with spirometry and multiple breath washout measures.

2 | Materials and Methods

This chapter describes the method developed for lung texture analysis and classification, which consists in a first algorithm for airways identification, which supports the second algorithm for lung parenchyma classification. The developed classifier is then applied to CT images of a group of patients with CF lung disease, which is described in the last paragraph of the present chapter. All the steps are implemented in python.

2.1. Image Processing Outline

The block diagram in Figure 2.1 shows the overall algorithm workflow.

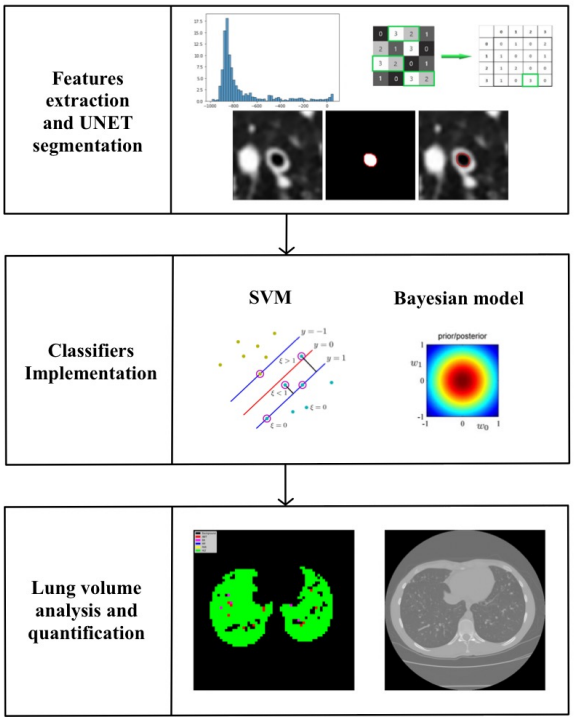


Figure 2.1: Overview of the algorithm pipeline. Images for SVM and the bayesian model are taken from [61].

Figure 2.2 describes in detail the creation of the dataset and the implementation of the classifiers.

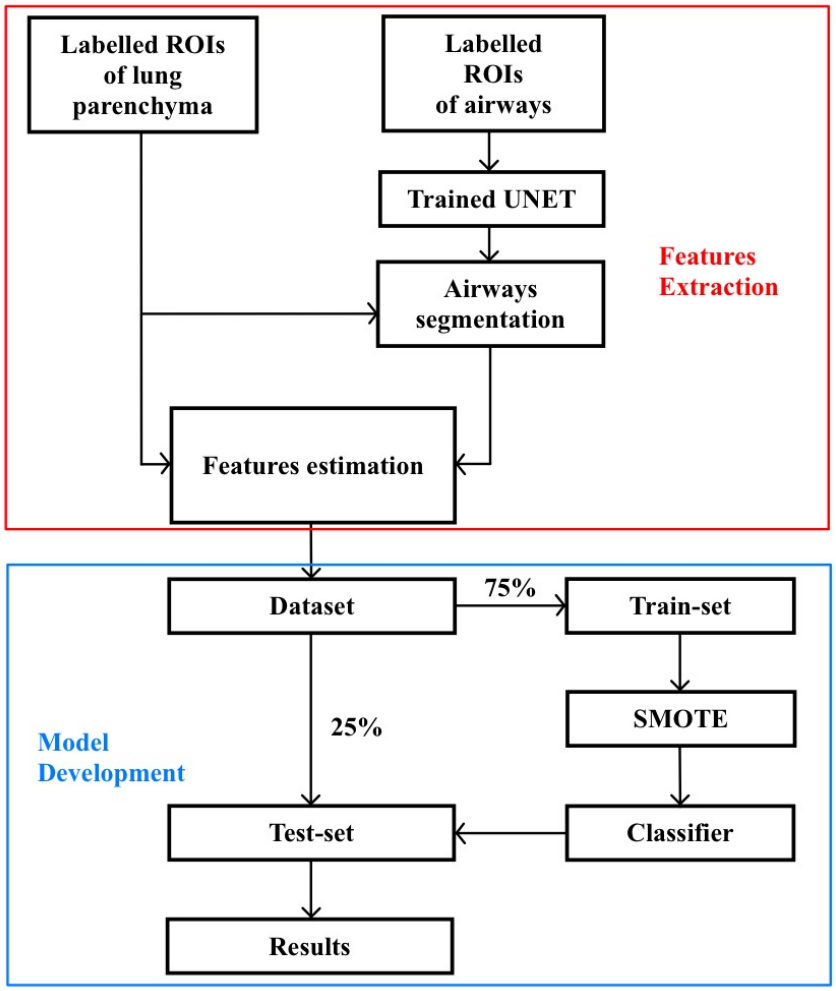


Figure 2.2: Features extracted from the ROIs are used to train and test the classifiers.

2.1.1. Dataset of labeled ROIs

Pathological and healthy regions of interest (ROI) were manually selected. According to the radiologist 4 classes were identified: airway wall-thickening (AWT), bronchiectasis (BR), mucous plugging (MP), and healthy (HLT). Respect to Brody score, the most used CT scoring system in CF lung disease, parenchymal abnormalities were not considered, as under-represented in our population. Squared ROIs were extracted from the centroid of manually drawn regions. ROIs sized 16x16 and 32x32 pixels were extracted, to find the optimal ROI size for the classification task. In Figure 2.3 there is an example of a ROI extracted from the original CT image.

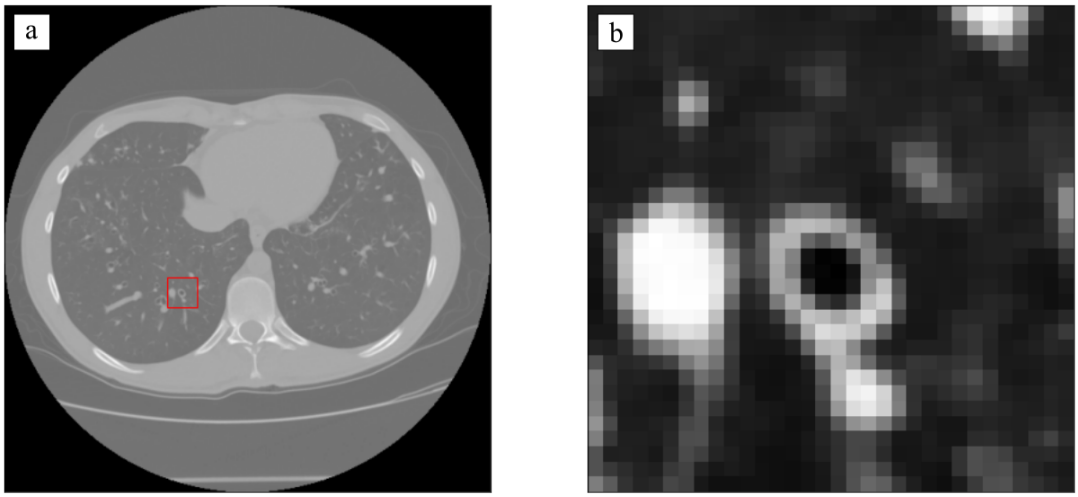


Figure 2.3: ***(a)*** HRCT slice on which the ROI is highlighted with a red square. ***(b)*** The same region is shown as the final image used for the analysis.

2.2. Features Estimation

For each ROI, a set of features have been calculated. In addition to the most common features present in literature, such as first and second order statistics, we implemented variables related to airways morphology and positioning within the lung, which have been demonstrated to be useful for pattern recognition in CF lung disease [58] (Figure 2.4). At the end of this section, all the features computed for each ROI are reported in Table 2.1.

2.2.1. Threshold Filtering

Thresholding is a common but fast tool to process the image. Thresholds are usually selected to highlight low or high attenuation patterns inside the lung. In this case, 4 ranges were chosen to underline the presence of different patterns: air-trapping, consolidation, ground-glass opacity (GGO) and high attenuation structures. Even if the name might be misleading, air-trapping doesn't refer only to the presence of air "trapped" inside of the lungs, but it also represents a hypo-perfused area [27]. Consolidation and GGO are high attenuation areas of the lung parenchyma, in which the vessels are still visible (GGO) or not (consolidation).

- Pixels $<$ -950 HU: for the evaluation of air-trapping inside of the lungs, as similar values can be found in literature [55].
- Pixels between -703 HU and -368 HU: for the evaluation of GGO [62].

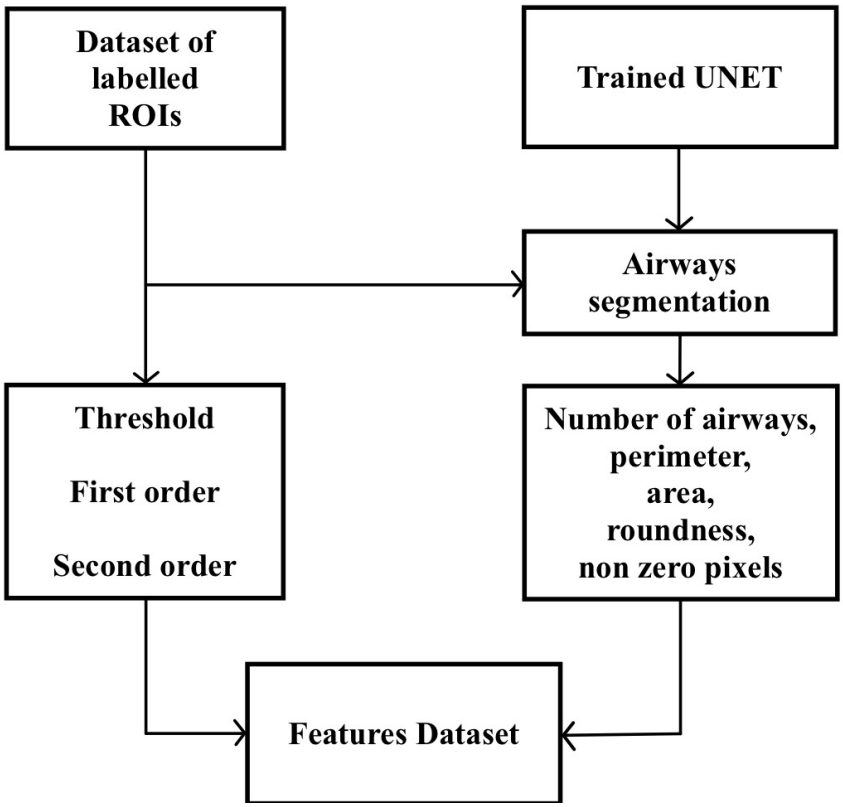


Figure 2.4: The features extraction process is composed of two parallel paths. In the first one, thresholds, first and second order statistics are computed directly on the ROIs. In the second path, airways are segmented with the help of a UNET and each ROI is represented with the number of airways, and informations relative to their shape and size.

- Pixels between -100 HU and 5 HU: for the evaluation of consolidation [62].
- Pixels $>$ -650: for the evaluation of all high attenuation structures inside of the lungs, similar to Chassagnon in 2018 [57].

The simplicity of this approach represents its strength and weakness as it requires a small computational time, but with a simple threshold lungs structures and pathological pattern cannot be distinguished (Figure 2.5). Moreover, the selection of fixed values for the thresholds does not consider the variability of CT images and that among subjects. Indeed, variable thresholds have shown a better correlation with FEV_1 than fixed thresholds [57]. After the filtering process, a quantitative evaluation was made by computing the number of non-zero pixels for each of the output images.

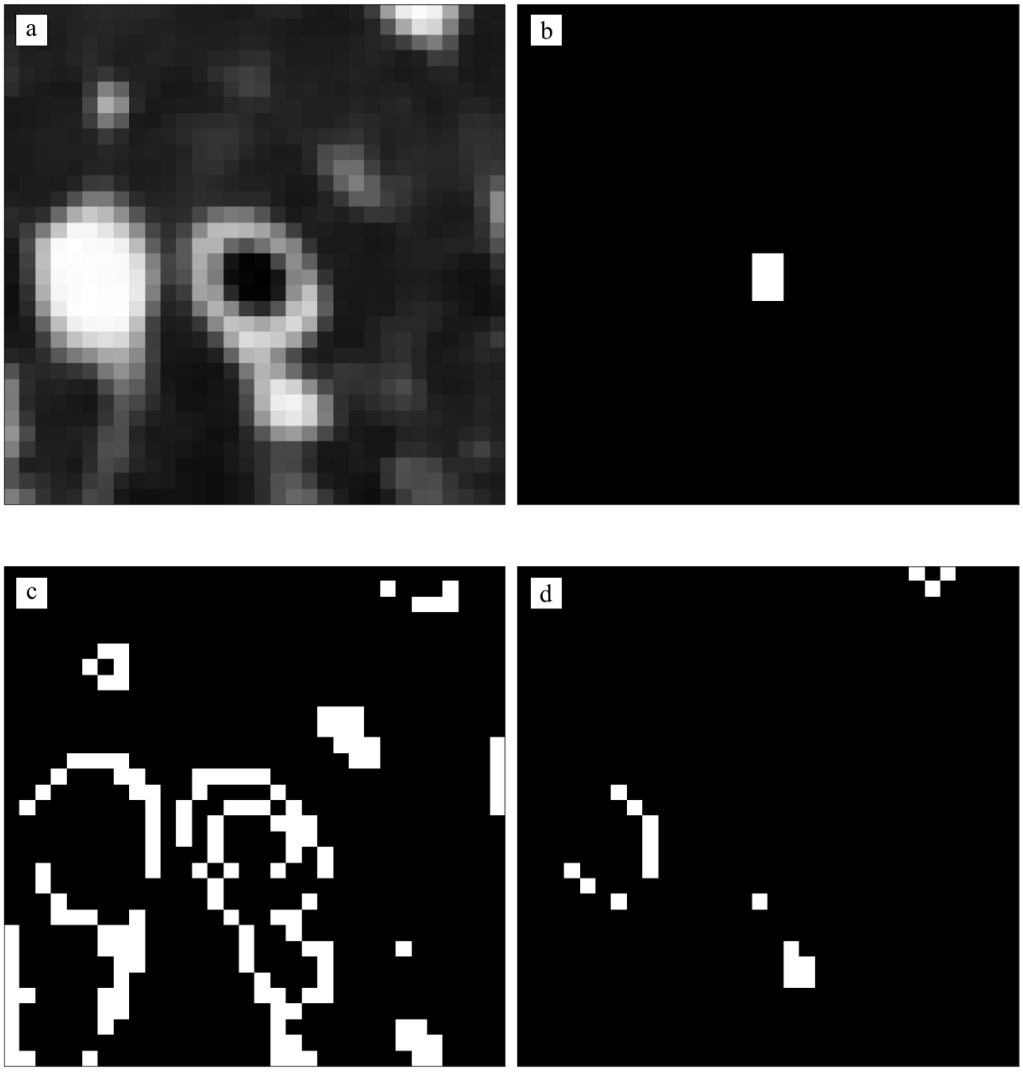


Figure 2.5: ***(a)*** Original image, before the application of threshold filters. ***(b)*** Mask with pixels identified as air-trapping. ***(c)*** Mask with pixels identified as GGO. ***(d)*** Mask with pixels identified as consolidation.

2.2.2. First Order Statistics

First order statistic parameters are computed on the image density histogram, which evaluates the frequency of the pixels at each density value. For each ROI the following parameters were calculated: mean, standard deviation, minimum, maximum, median, mode, skewness and kurtosis. Nevertheless, the parameters do not account for the spatial relationship between the pixels [50].

A representative ROI histogram with the corresponding features is reported in Figure 2.6.

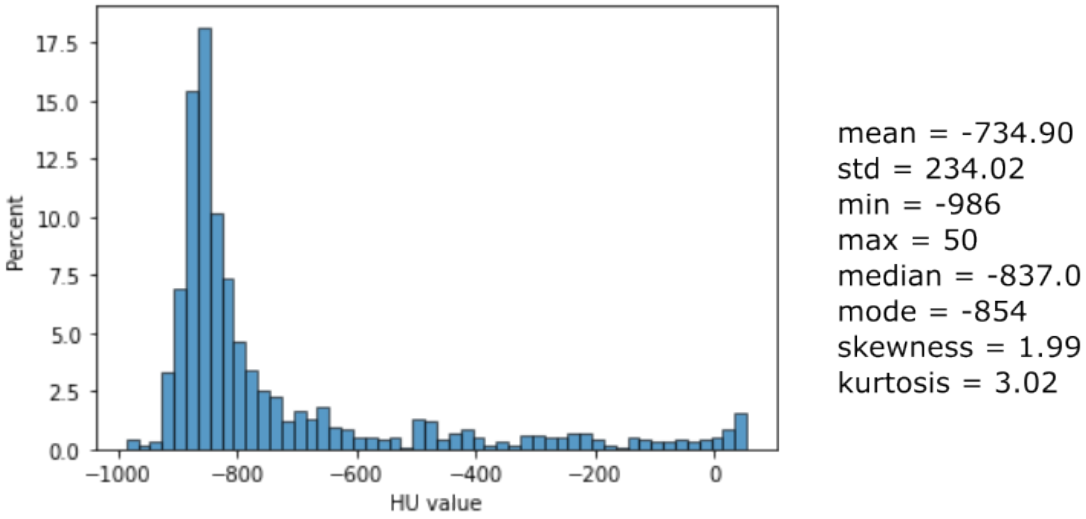


Figure 2.6: ROI density histogram and his first order statistics parameters.

2.2.3. Second Order Statistics

The gray-level co-occurrence matrix (GLCM) was calculated to derive second order statistic parameters. Prior to the computation of the GLCM matrix, input images are quantized in a finite number of level, 16 discrete levels ranging from 0 to 15, and each ROI was processed as shown Figure 2.7.

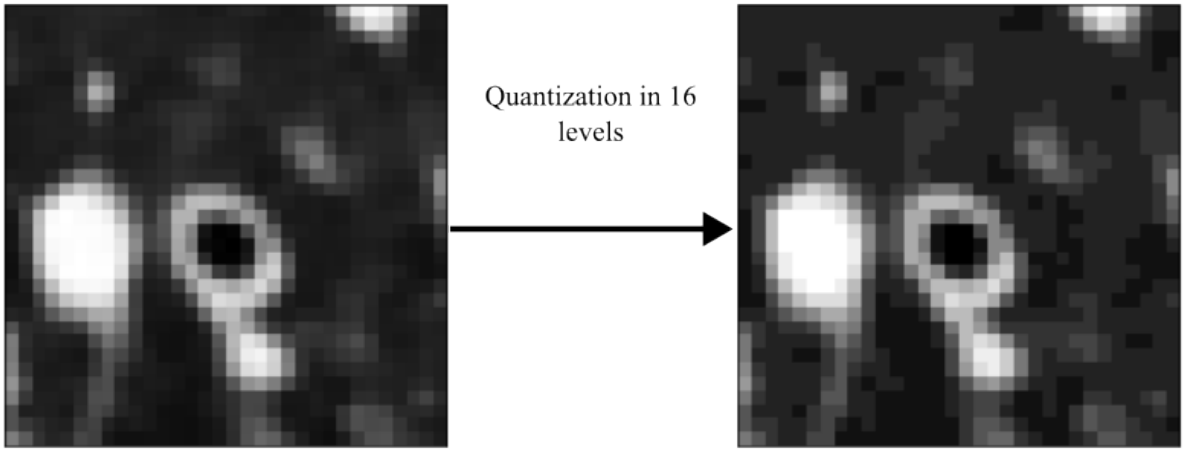


Figure 2.7: Quantization of a ROI into 16 discrete levels.

GLCM counts the number of times a pair of pixels of levels i and j appear at a given distance d and angle θ. In order to completely explore a 2D neighborhood, GLCMs were computed with one distance d equal to one pixel and four directions θ: 0°, 45°, 90°, 135°.

In Figure 2.8 is represented the operations to compute a GLCM for d = 1 and $\theta = 0°$, of a 4 levels image.

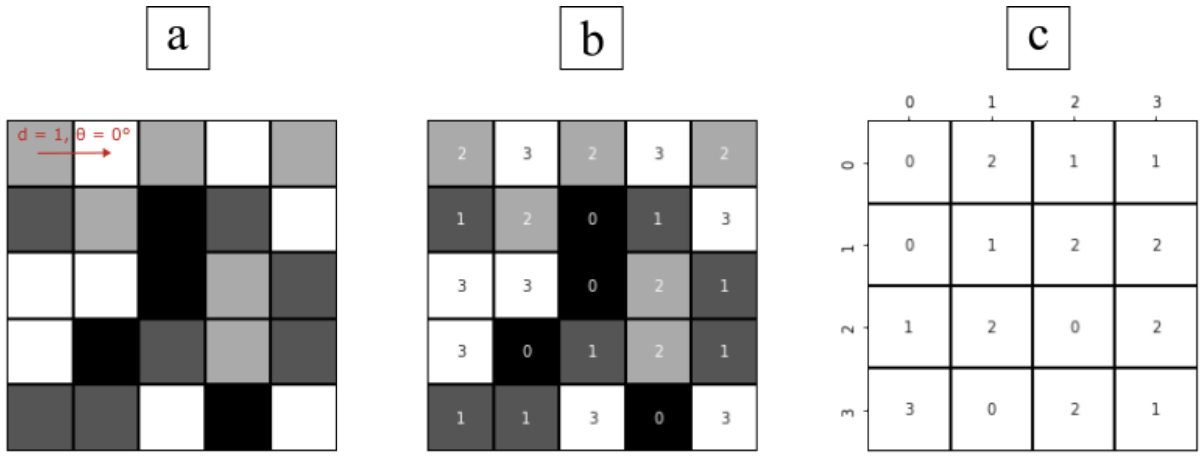


Figure 2.8: ***(a)*** Example image, in which the direction and angle are shown. ***(b)*** Example image, with annotation of the value for each pixel. ***(c)*** Final GLCM.

Then, from each GLCM, the following parameters were computed:

$$contrast : \sum_{i,j=0}^{levels-1} P_{i,j}(i-j)^2$$

$$dissimilarity : \sum_{i,j=0}^{levels-1} P_{i,j}|i-j|$$

$$homogeneity : \sum_{i,j=0}^{levels-1} \frac{P_{i,j}}{1+(i-j)^2}$$

$$ASM : \sum_{i,j=0}^{levels-1} P_{i,j}^2$$

$$energy : \sqrt{ASM}$$

$$correlation : \sum_{i,j=0}^{levels-1} P_{i,j} \left[\frac{(i-\mu_i)(j-\mu_j)}{\sqrt{(\sigma_i^2)(\sigma_j^2)}} \right]$$

where $P_{i,j}$ is the pixel at the entry (i,j) of the normalized GLCM and $\mu_x, \mu_y, \sigma_x, \sigma_y$ are the mean and standard deviation of $p_x(i) = \sum_{j=1}^{levels} P_{i,j}$, and $p_y(j) = \sum_{i=1}^{levels} P_{i,j}$ [46].

2.2.4. ROI Position Compared to Lung Contour

As anticipated in section 1.4, the presence of airways in the most peripheral part of the lungs are indicative of the presence of bronchiectasis. Therefore, a specific feature was developed, computing the minimum distance between the center of each ROI and the lung contours. To this aim, lungs contours are identified, using the "find_contour" function which is implemented in the skimage library of python [63] (Figure 2.9). The function finds contours searching for constant values inside the image using the marching squares method. The marching squares algorithm uses the same approach as marching cubic algorithm [64], but just in two dimensions. As the function returns all the contours inside of the image, they are filtered based their closure and size, to retrieve only the lungs. Once the lungs' contours are found, the minimum distance from the edges of the lungs to the center of a ROI can be easily obtained.

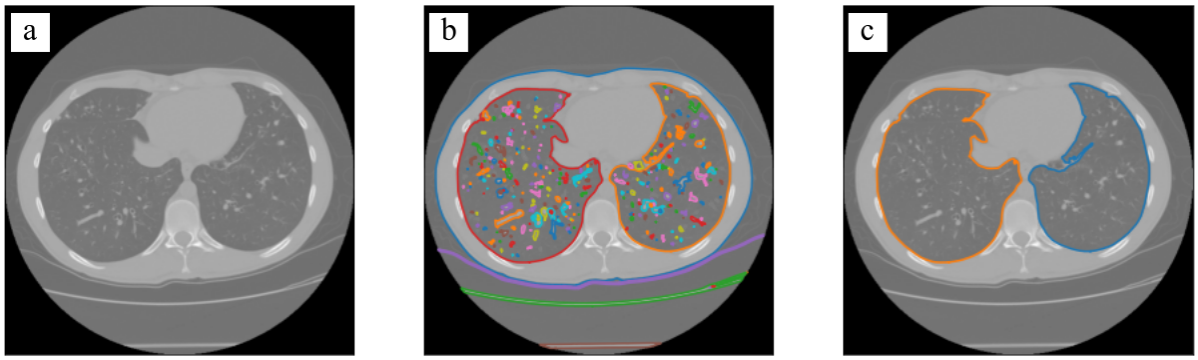


Figure 2.9: Identification of lungs contours. ***(a)*** Input image. ***(b)*** All the contours around a defined level are found. ***(c)*** Then they are filtered by taking into account if they are a closed loop and their size, in the end only the two lungs remain.

2.2.5. Airways Features

Airways Detection

The correct identification of the airways is of utmost importance in the evaluation of pathological pattern in CF lung disease, especially for the identification of airway wall-thickening and bronchiectasis.

I implemented two different approaches for airways identification.

The first approach was similar to the one previously described for the delineation of lung contours. All the contours at a defined HU level were found and then filtered to retrieve the airways. An airway was accepted if the contour was close and if the object enclosed

a low density area (airway lumen) and was completely surrounded by a high attenuating structure (airway wall). The process is outlined in in Figure 2.10.

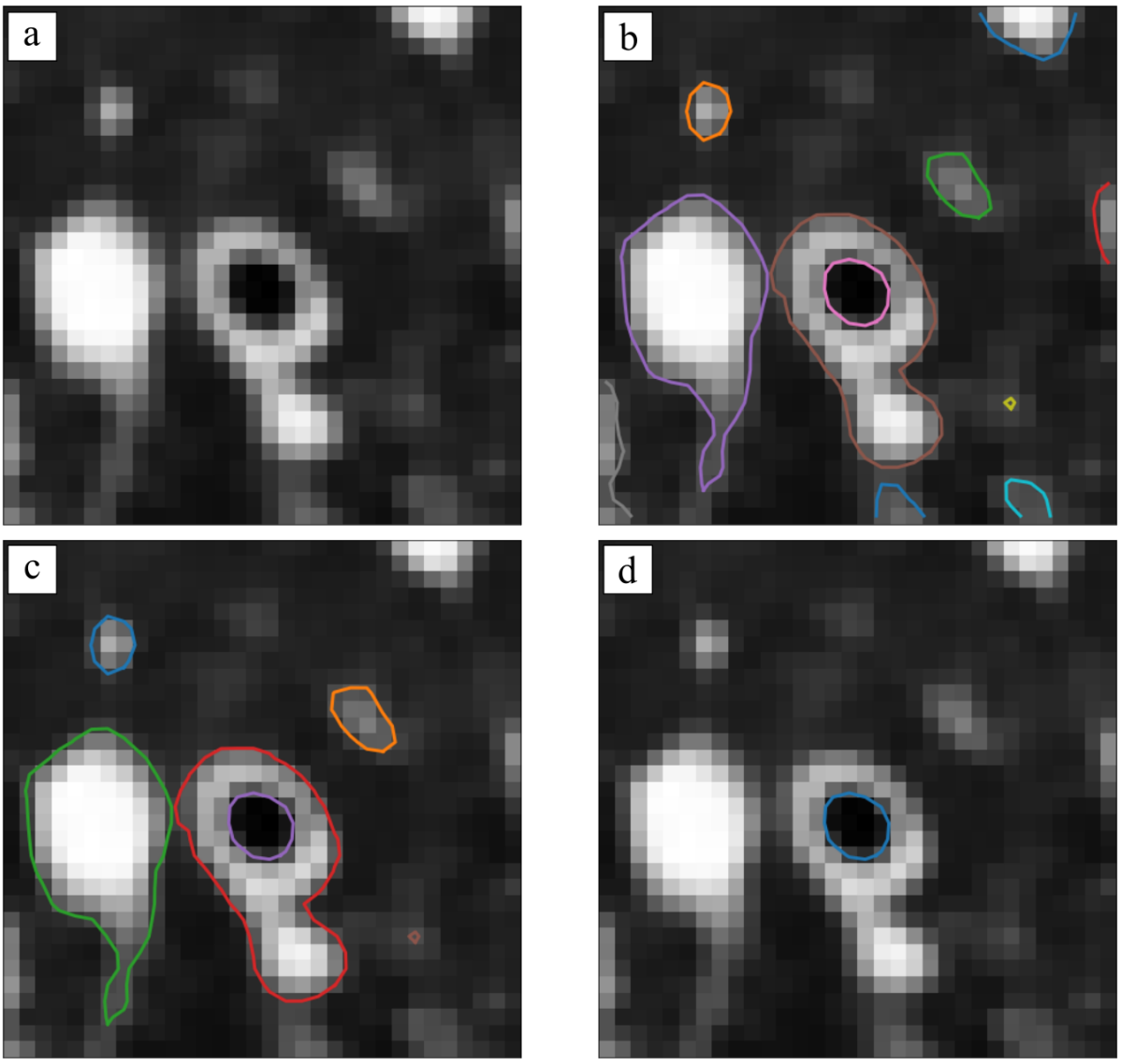


Figure 2.10: ***(a)*** The process begins with the original image. ***(b)*** All the contours, open and closed, are found around -700 HU. ***(c)*** A first check is made on the contours and only the ones closed are kept. ***(d)*** Then, the mean value for the region defined by each contour is computed, and if it is below a defined threshold (-750 HU) and it is surrounded by high attenuation structures, the contour is accepted as an airway.

This approach had limitations, as not only each airway had its own intensity level to correctly identify the contour, but also the parameters used to identify the lumen and the surrounding wall had to be changed. This first approach cannot be extended to a larger scale, thus an automatic method based on deep-learning was implemented.

The chosen architecture was the UNET, due to its proven ability to segment biomedical images and its structure is shown in Figure 2.11.

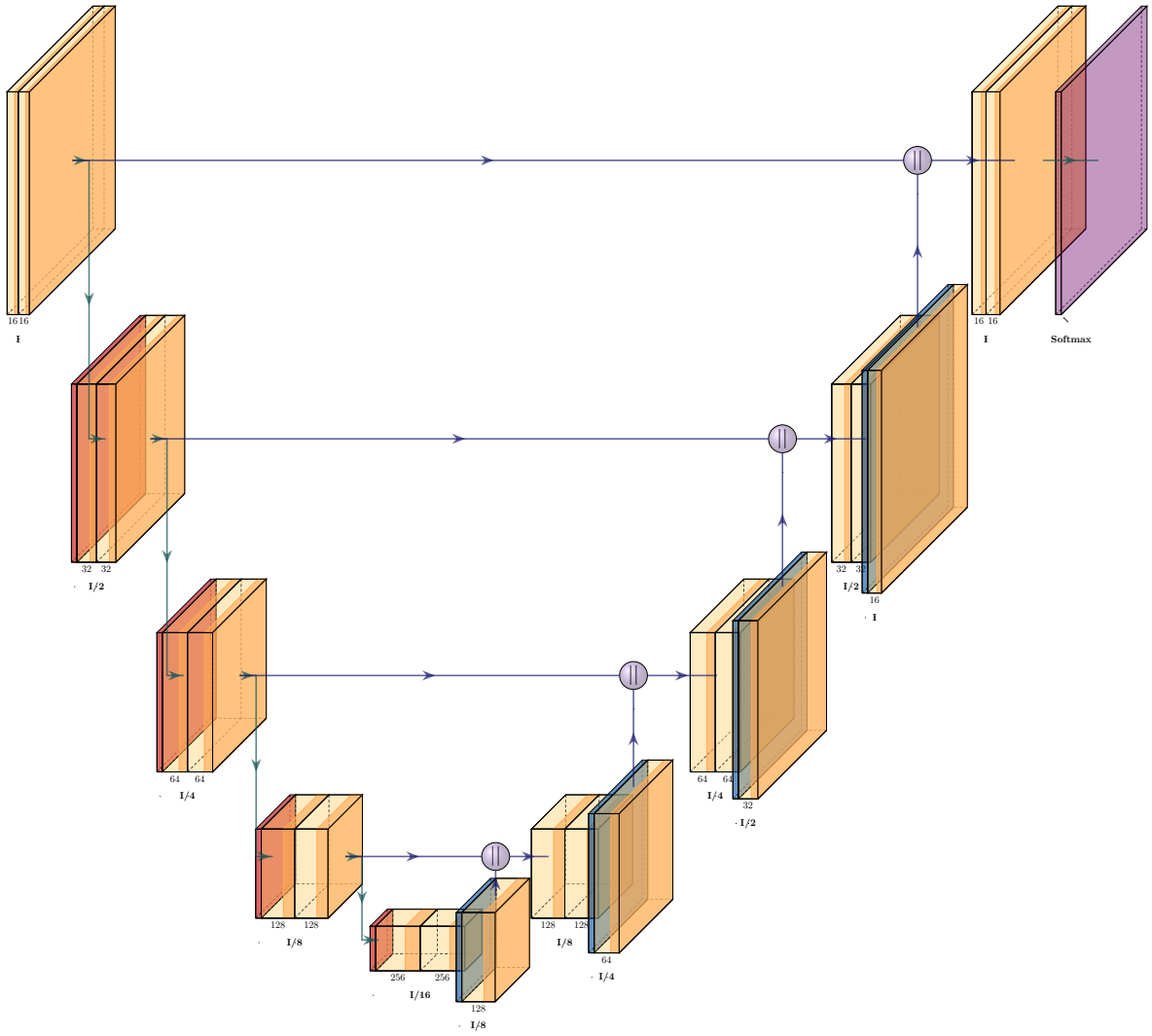


Figure 2.11: Architecture of the UNET.

The UNET architecture is composed of two parts: the contracting path and the expansive path [65]. In the contracting path, two convolutions with kernel of size 3x3 are followed by rectified linear unit (ReLU) and max pooling. In the expansive path, the input features map is first upsampled and the number of features is divided by two with a 2x2 convolution. Then the features map is concatenated with the corresponding features of the contracting path and two 3x3 convolution are applied. The starting number of feature map selected for the UNET was 16. A lower value caused a worsening of the performance both on validation and test set, while a higher value increased the risk of over-fitting without providing any benefit in the performance of the model.

The dataset used for the training was manually built using the approach based on the contour identification and was composed by 10373 images, which could contain or not airways. Before training, the images were upscaled to three time the original size, to both increase the airway size, that in some cases could be close to image resolution, and to allow a smoother transition between adjacent pixels, which improve contour identification. The loss function selected was the intersection over union which is defined as:

$$\text{IoU loss} = 1 - \frac{\text{Area of intersection}}{\text{Area of union}}$$

and evaluates the performance of the model by comparing the agreement of the intersection and union between the predicted image and the target mask.

Once the model was trained, it was used to propose airways candidates, which still undergo a validation process in order to be accepted. The candidate contour was identified in the prediction image. Then, on the original image, the contour had to satisfy the same requisites of the previous method. A candidate was accepted if the contour identified a low-density area, completely enclosed by a higher density structure. The advantage of this method is to automatically identify possible airways without constantly tuning the parameters of the previous approach. This is possible because the contours are found in the prediction and then used to extract information from the original image as shown in Figure 2.12.

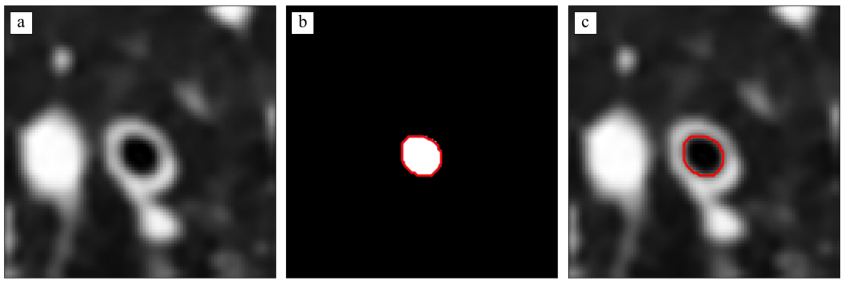


Figure 2.12: ***(a)*** The input image is first scaled to three times the original size. ***(b)*** Then, the prediction is computed using the UNET and the candidate contours are identified. ***(c)*** The lumen intensity and the presence of a higher attenuation structure around it are checked before validating the candidate.

Airways Features

When a ROI was analyzed, for each segmented airway, area, perimeter and roundness were extracted. Thus, each ROI was characterized by the number of airways, their average area, average perimeter, average roundess 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. Since these values are computed on the resized image, they are first rescaled for the dilation factor and then converted in millimeters. The conversion from pixels to millimeters was needed because the patients' CT scans didn't have the same spacing in the horizontal plane, thus the real dimension of the airways did not correspond to the number of pixels.

Features

Method	Feature
Threshold	Air-trapping Consolidation GGO High attenuation structures
First Order	mean standard deviation min max mode median kurtosis skewness
GLCM	Contrast Dissimilarity Homogeneity ASM Energy Correlation
Airways and Lung	Number of airways Average perimeter Average area Average roundness Presence of Airways Distance from edges

Table 2.1: Summary of all the features extracted for each ROI.

2.3. Classification

In this section is described the composition of the dataset, the features selection process and the classifier implementation.

2.3.1. Dataset

The final dataset was composed of 472 ROIs divided in 4 classes as follows: 79 airway wall thickening, 142 bronchiectasis, 86 mucous plugging, 165 healthy regions. In Table 2.2 is reported the number of samples per class in the dataset.

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

Table 2.2: Composition of the dataset.

2.3.2. Classifier Implementation

In this subsection a short description of the two classifiers implemented, i.e. Bayesian classifier and Support Vector Machine (SVM), is provided.

Bayesian Classifier

The aim of a Bayesian approach is to learn the posterior probability over the model parameters w, from the data and a prior probability distribution.

In the "naïve" bayesian approach it holds the hypothesis that every pair of features are independent given the class. The inference process is based on the Bayes theorem:

$$p(\text{parameters} \mid \text{data}) = \frac{p(\text{data} \mid \text{parameters}) \cdot p(\text{parameters})}{p(\text{data})} \propto \text{likelihood} \cdot \text{prior}$$

where $p(\text{parameters} \mid \text{data})$ is the posterior over the training data. The parameters that will be chosen are the mode of the posterior probability distribution, also known as maximum a posteriori (MAP).

For our work we used a Gaussian Naive Bayes algorithm where the likelihood is assumed to be:

$$p(x_i \mid y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} exp\left(\frac{(x_i - \mu_y)^2}{2\sigma_y^2}\right)$$

In Figure 2.13 , an example of the sequential Bayesian process is shown.

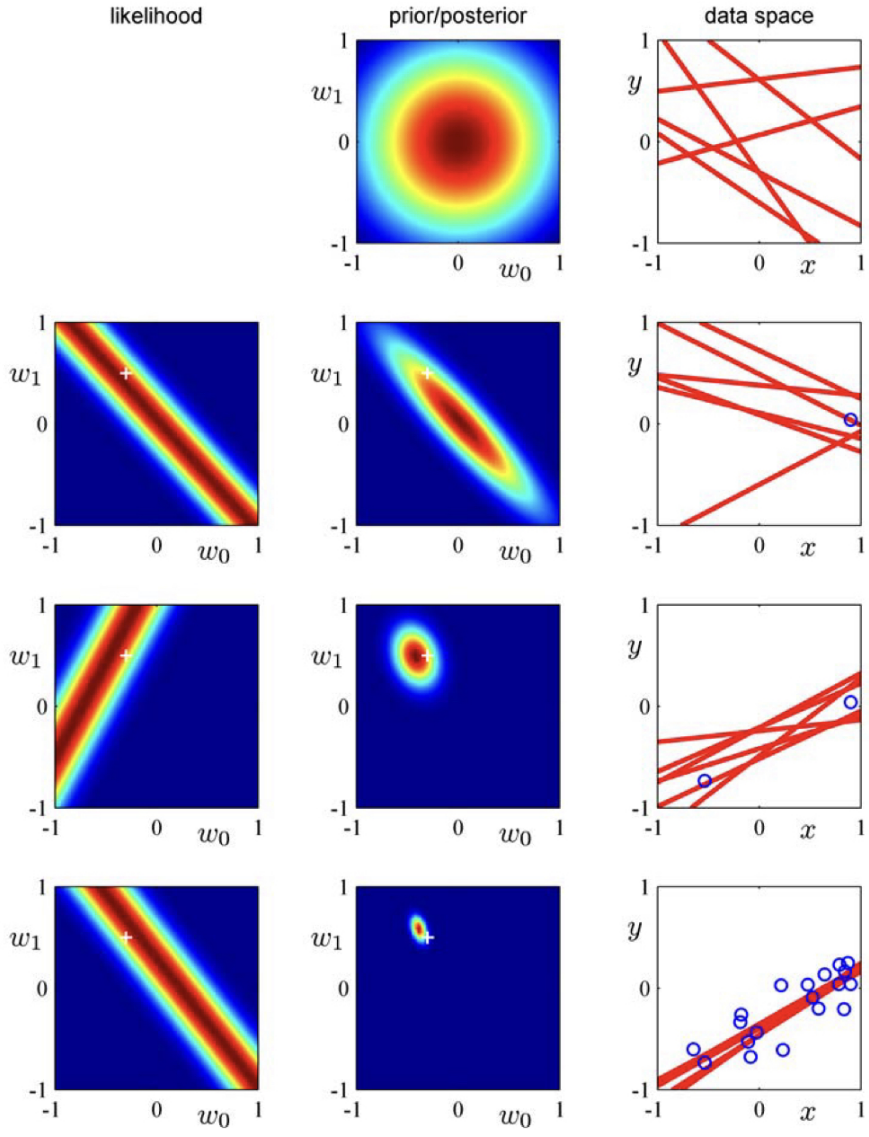


Figure 2.13: The figure shows the Bayesian sequential learning. As the model experience data the posterior is tuned, and then it is used as prior in the following iteration [61].

Support Vector Machine (SVM)

SVM is a classification method that tries to maximize the margin between the decision boundary and the closest data points. Thus, the decision boundary is defined only by a subset of the data points, which are known as support vectors. In Figure 2.14, an example of SVM for binary classification is shown.

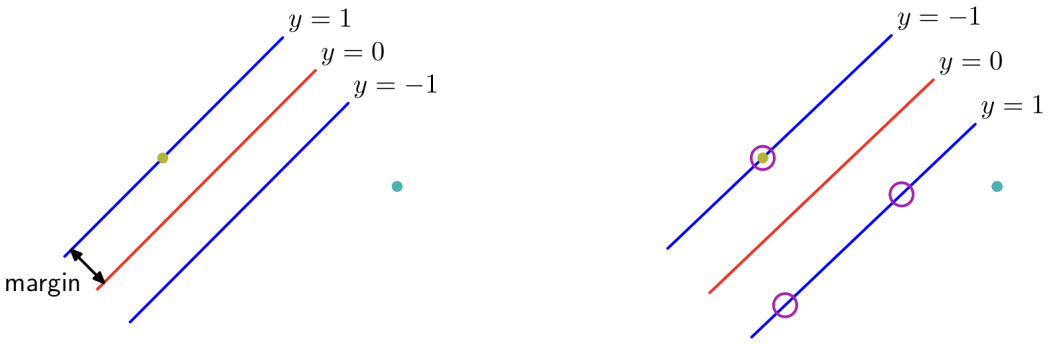


Figure 2.14: Only the closest points contribute to the definition of the decision boundary [61].

Since real problem are not often linearly separable, the input $\bar{x}$ is usually mapped in a higher dimensional space with the help of features functions $\Phi(\bar{x})$ or kernel trick. In a binary classification task, the target class for a sample $\bar{x}$ computed by SVM is:

$$y(\bar{x}) = sign(w^T \Phi(\bar{x}) + b)$$

where w and b are the weights and bias and are found by maximizing the margin in the train dataset. But even in an higher feature space a perfect solution might not be found, for these reason soft margin ξ_n are introduced to give some compliance to the model. With the introduction of soft-margins, data points are allowed to be inside of the margin and even to be missclassified. Thus the minimization function becomes:

$$\frac{1}{2}||w||_2^2 + C\sum_{n=1}^{N}\xi_n$$

subjected to:

$$\begin{cases} y_n(w^T\Phi(x_n)+b) \geq 1-\xi_n, \text{for all n} \\ \xi_n \geq 0, \text{for all n} \end{cases}$$

C is an hyperparameter that controls the trade-off between margin and error. SVM solves binary classification problems, but it can be extended to multiple-classes by exploting strategies as One vs One or One vs All. In Figure 2.15, the effect of soft-margins addition can be seen.

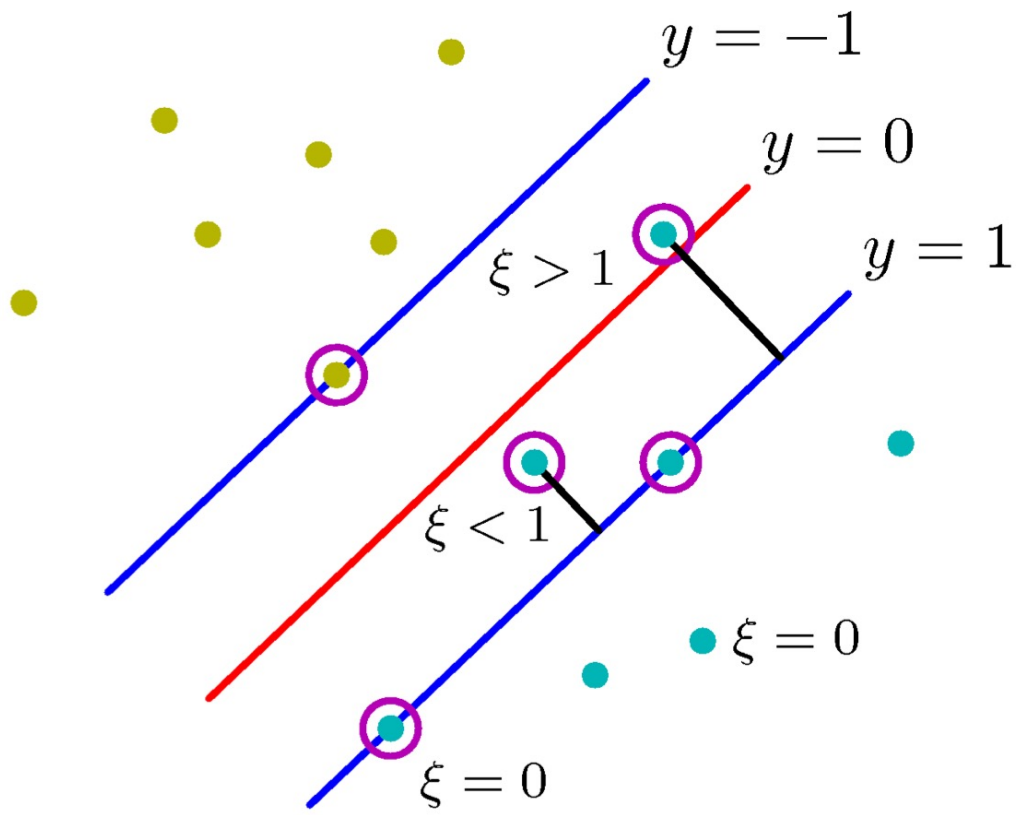


Figure 2.15: with soft-margins, points are allowed inside of the margin and they can even be missclassified [61].

2.3.3. Cross-Validation

In a machine learning approach, the selection of hyper-parameters is an important task in order to optimize the performance of a model. At first glance, a good strategy would be to just train the model and tweak the hyper-parameters in order to maximize the metrics used on the test set. This approach risk to increase over-fitting because the hyper-parameters are tuned directly on the test set and thus the performance of the model is not independent from test data. A solution is to divide the dataset in 3 sets: train, validation and test set. In this case, the validation set is used to tune the model hyper-parameters without introducing the influence of the test set. This solution is unfeasible in problems where the number of samples is limited and other strategies must be used, as cross-validation. In this setting, a validation set is not required as the training set is split in k subsets: k-1 are used for the training phase and the last one is used for the validation. The procedure is done for each of the k-folds, allowing to validate the model without the influence of the test set. In Figure 2.16 is shown a training set split in five folds, four will be used for training and one for validation. In this work, the train set was divided in five "stratified" folds. The term "stratify" means that the folds are created by maintaining the same percentage of samples for each class [66].

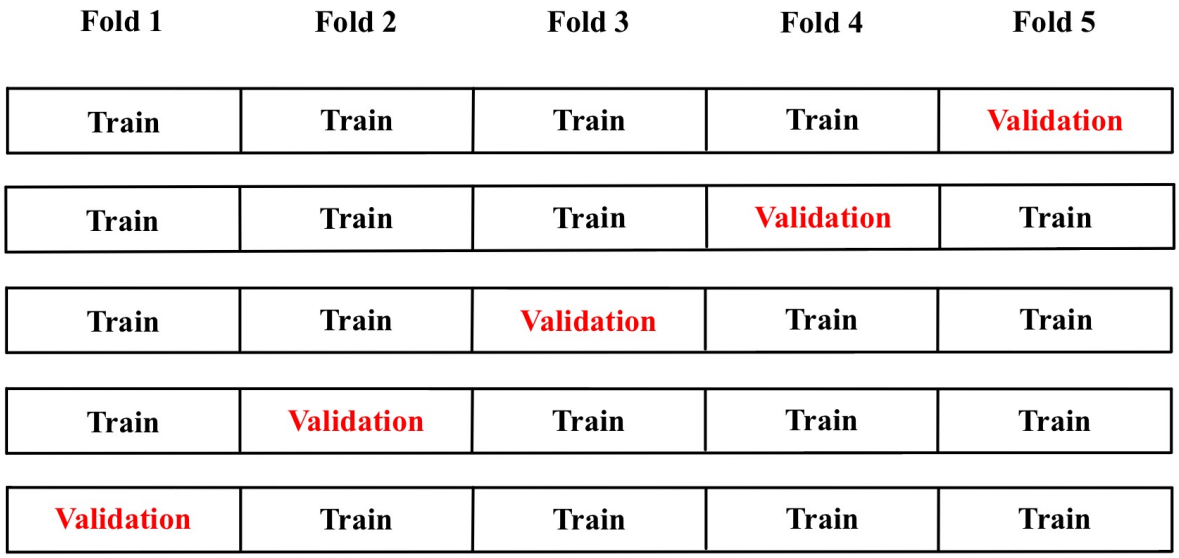


Figure 2.16: Example of k-fold cross-validation with k = 5.

2.3.4. Features Selection

Feature selection is used to minimize the amount of variance into the model by reducing the number of features on which is trained. Other techniques that can achieve the same scope are dimensionality reduction, which reduces the dimension of the input space, or regularization, to keep the parameters of the model close to zero. In feature selection, the features that are deemed redundant or irrelevant are removed from the training set with different techniques. The first approach used was a forward sequential feature selection, but really few features were discarded, and the trained model showed strong over-fitting on real images.

Then, it was tried a strategy based on correlation of the data, using Pearson correlation coefficient:

$$\rho_{x,y} = \frac{cov(x,y)}{\sigma_x \sigma_y}$$

The correlation coefficient was computed for each pair of features and highly correlated features (ρ >0.9) are removed. Then, the correlation coefficient was computed between each feature and the target class, retaining only features with a good correlation (ρ > 0.4). From 24 features, only 8 were selected.

In the table 2.3, the final set of features is shown.

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

Table 2.3: Features after the features selection process.

2.3.5. SMOTE

A balanced dataset has the same number of samples for each class, while an imbalanced dataset has a non-homogeneous distribution of samples among the classes. The classes which compose most of the dataset are referred to as majority classes, while the others as minority classes. A model trained on an imbalanced dataset may be biased, due to a high degree of accuracy for the majority classes and a low accuracy for the minority classes.

Imbalanced datasets are a common problem for machine learning applied to medical tasks as usually there are more samples for the healthy class than for the pathological ones [67]. As it can be seen in subsection 2.3.1, our dataset is imbalanced and thus Synthetic Minority Oversampling Technique (SMOTE) was applied [68]. SMOTE is a technique used to deal with imbalanced problems as it generates artificial samples by exploiting similarities in the feature space in the minority classes.

To create a new synthetic element, a sample $\hat{x}$ belonging to a minority class is chosen. Then, one of the k-nearest neighbour of the same class x is randomly selected and a new synthetic sample is generated as follows:

$$x_{\text{new}} = \hat{x} + (x - \hat{x}) \cdot \delta$$

where $\delta \in [0, 1]$ is a random number.

The Figure 2.17 shows the scatterplot of two variables in both the original and the over-sampled dataset. In the images SMOTE is applied to the whole dataset only for a better representation of the data, but during the classification it will be used only on the training set.

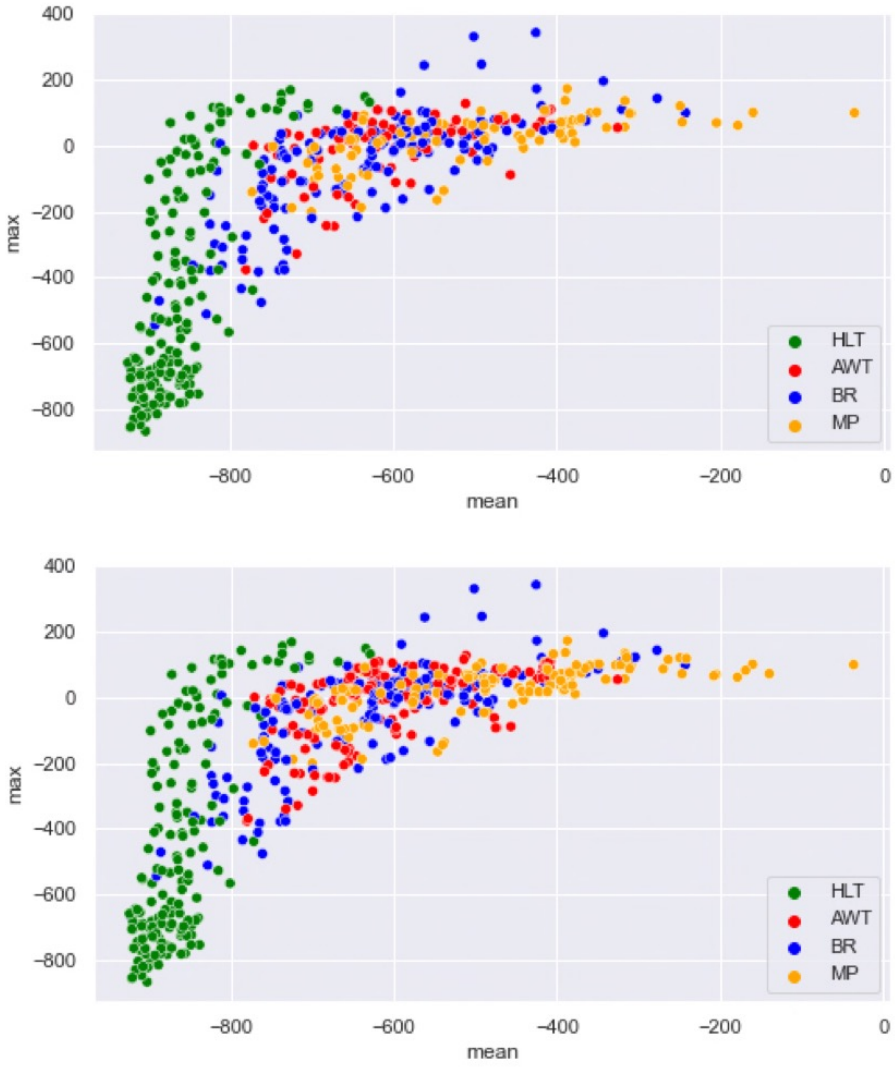


Figure 2.17: Scatterplot of mean vs standard deviation before and after SMOTE.

SMOTE generates new samples without considering the presence of samples from other classes. This process can lead to overlaps among different classes, and this effect can be noted in Figure 2.17. For this reason, more advanced algorithms, such as borderline-SMOTE [69] or ADASYN [70], were proposed, but the selection of the best up-sampling technique was not in the scope of the thesis.

2.3.6. Metrics

The performance of the classifiers was measured with the following metrics: sensitivity, precision, specificity and f1 score weighted for the classes.

Sensitivity measures the ability of a model to correctly classify positive samples in the positive class.

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

Precision measures the percentage of samples that a model predicts as positive and really belong to the positive class.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

Specificity measures the ability of a model to correctly classify negative samples in the negative class.

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}$$

The f1 score is the harmonic mean of precision and recall. It combines the information coming from the two metrics and thus allows an easier comparison between two models.

$$\text{f1} = \frac{2 \cdot \text{Precision} \cdot \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}}$$

2.4. Volume Processing Outline

In this section the process for volume evaluation is described (Figure 2.18). First, the images are pre-processed by removing vessels with the Frangi's filter. Then, lungs are segmented and each pixel is classified according to the trained classifier.

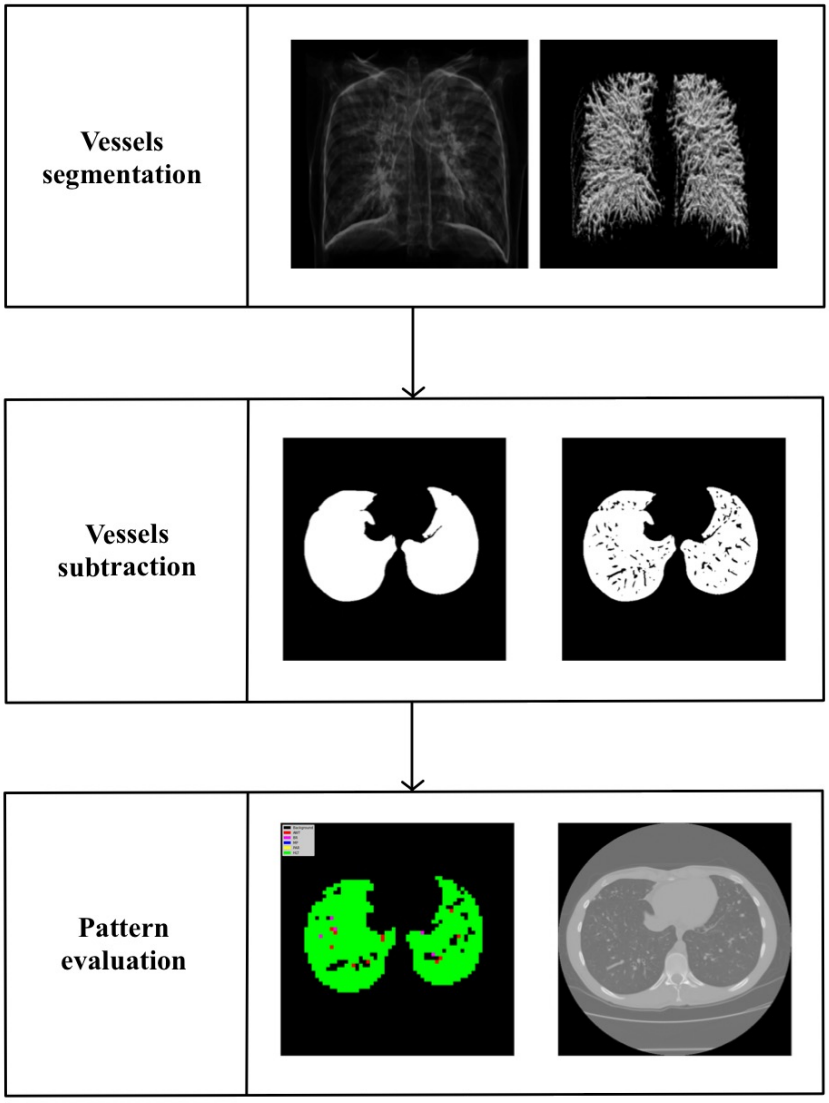


Figure 2.18: Evaluation workflow. First the vessels are segmented, then they are removed from the original image and the ROIs which have at least 80% of the pixels belonging to the lung are chosen. In the end, the accepted ROI are classified.

2.4.1. Vessels' segmentation

Vessels appear in CT images as high attenuation structures, which could be confused with consolidation or mucous plugging. In order to prevent their impact during the evaluation process, they are removed from the image before classification. For the vessels' segmentation Frangi's filter is used [71]. The algorithm is based on the computation of the hessian matrix of the image in the 3D space and enhance rounded, tubular or planar structures based on the eigenvalues. In Figure-2.19, it is shown an example of the final segmentation.

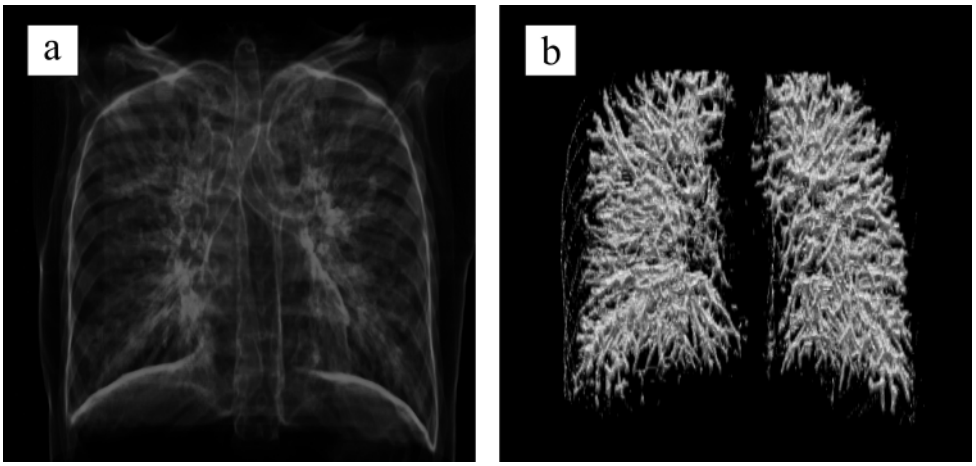


Figure 2.19: ***(a)*** Original image. ***(b)*** Highlight of the vessels found inside of it, please note that the main branches of the airways have been removed from the segmentation because they weren't useful for the scope of the work.

2.4.2. Volume Evaluation

The lungs were segmented as described in subsection 2.2.4. Then, for each slice, a grid was constructed starting from the bounding box of the lung. Only the ROI with at least 80% of pixels being lung pixels were accepted. The window slides across the grid cells and classifies the pixels inside it (Figure 2.20). Hence, while a bigger window slides across the image, only the central pixels corresponding to the grid cells are classified at each step. Then, the window moves on the next cell and the process is repeated until the lungs are fully covered.

The grid was built with a spacing of 8 pixels, thus creating cells of 8x8 pixels. This value was selected as a trade-off between computational time and output resolution. A bigger value allowed to reduce the computational time but had a worse resolution on the output image. The difference is shown in Figure 2.21.

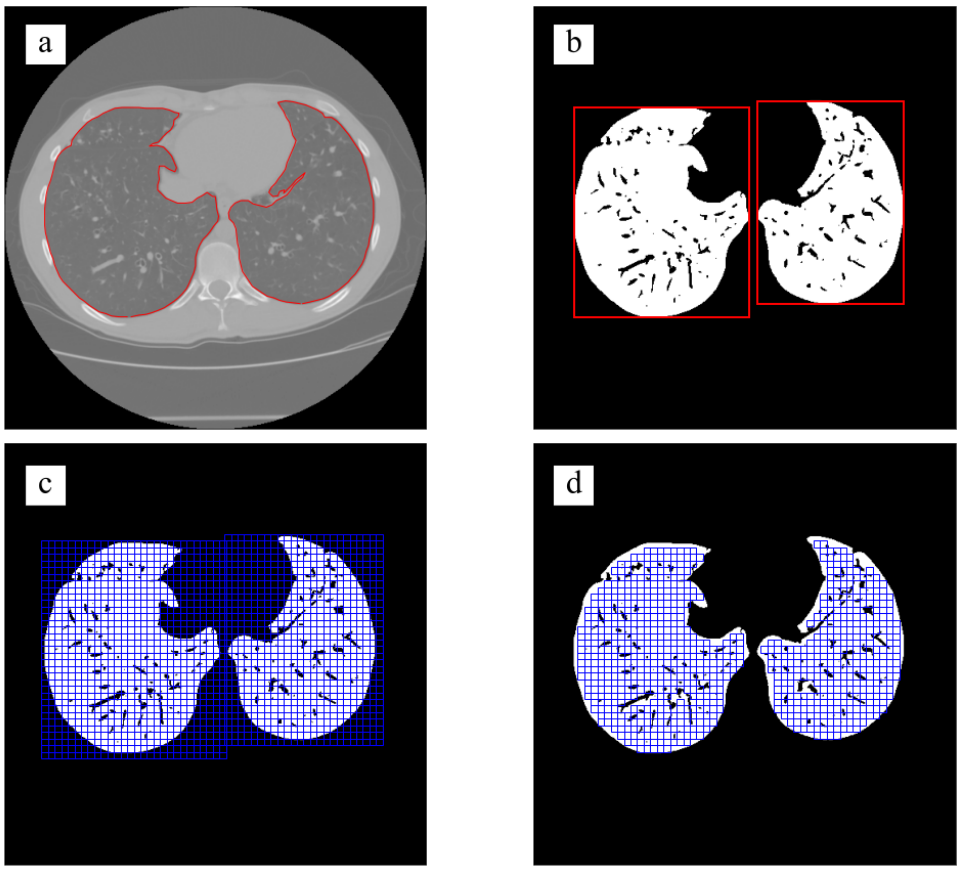


Figure 2.20: The image shows the process that identifies the lung regions to be classified. ***(a)*** First the lungs' contours are identified. ***(b)*** Then the bounding box is derived. ***(c)*** A grid is created with cells of desired size. ***(d)*** Only the cells containing at least 80% of lung pixels are retained.

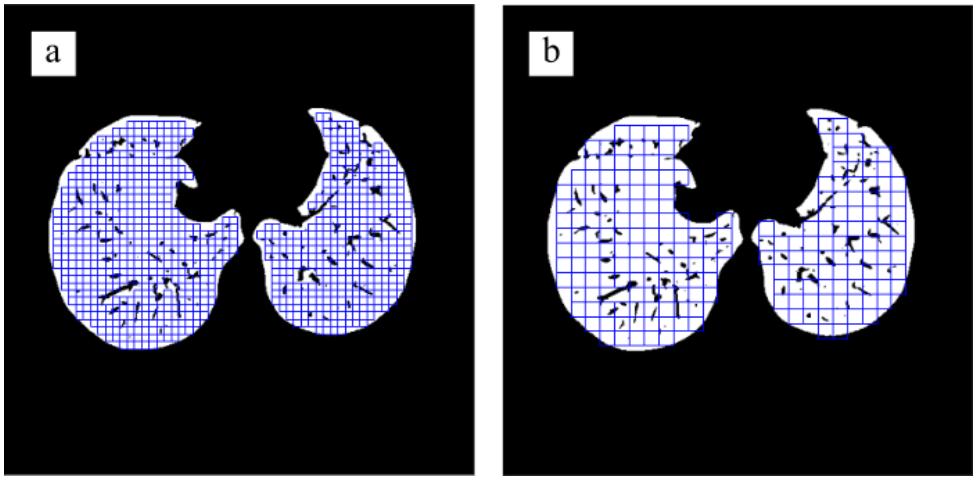


Figure 2.21: The same lung is divided in region sized 8x8 ***(a)*** pixels and 16x16 ***(b)***.

2.4.3. Quantitative Evaluation

The segmented vessels were subtracted from the HRCT volume and each disease class (AWT, BR, MP and HLT) was quantified on the overall lung volume. Each class was expressed as percent volume (AWT%, BR%, MP%, HLT%). As the condition of the airways is a key point in cystic fibrosis lung disease, also the airways volume was evaluated as a percentage over the lung volume (vol_a%).

2.4.4. Study Population

A group of patients with cystic fibrosis (CF) lung disease was studied. Diagnosis of CF was confirmed by a sweat chloride level >60 mM and/or two CF-causing mutations in the CFTR gene according to established diagnostic criteria. All patients underwent chest CT, spirometry and MBW test as part of their annual surveillance programme at the Lombardia Region CF Centre, University of Milan, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy. Spirometry was performed on the same day as CT and included the measurement of FEV1, forced vital capacity (FVC) and FEF25–75%. Values were expressed as a percentage of predicted values according to the prediction equations of the Global Lung Function Initiative. Spirometry was performed in compliance with American Thoracic Society/European Respiratory Society (ATS/ERS) guidelines [72]. Multiple breath washout was performed within 15 days of CT. The commercial nitrogen-MBW equipment ExhalyzerD and associated software package (Eco-Medics AG, Duernten, Switzerland) were used for data acquisition, storage and analysis in compliance with the standard operating procedure [73]. PFTs of patients are reported in Table 2.4.

PFTs	**Median (25th - 75th percentile)**
FEV_1 (L)	2,90 (1,76 - 3,44)
%FEV_1	70,70 (54,70 - 82,15)
FVC (L)	4,48 (3,00 - 5,00)
%FVC	86,90 (76,40 - 103,25)
$FEF_{25\%-75\%}$ (L)	1,73 (0,80 - 2,04)
%$FEF_{25\%-75\%}$	36,40 (22,7 - 46,9)
LCI 2,5% norm	15,17 (13,36 - 17,11)

Table 2.4: PFTs expressed as median and 25th-75th percentile.

3 | Results

In this chapter the results of this work are shown. First, the performance of the UNET for airways segmentation is described, then the comparison between the SVM and the Gaussian Naive Bayes (GNB) classifiers is reported. The performance of the best model on the training set was tested on ROIs sized 16x16 and 32x32 pixels, to find the best size for the classification task. In the end the best model is used to quantify the different lung patterns over the whole lung volume and the results are compared to radiological scores and PFTs.

3.1. UNET Evaluation

The "airways dataset" included 19 patients, for a total of 10373 ROIs. The dataset was split into training, validation and test set based on patient ID, in order to have a better generalization. 17 patients (9305 ROIs) were used for the training, 1 (490 ROIs) for the validation, and 1 (578 ROIs) for the test set. At the end of the training, the model presented a validation loss of 0.315 and a test loss of 0.305. Since the test loss does not have relevance by itself, the model performance was manually checked by testing it on other 100 sample ROIs extracted randomly from the "lung parenchyma dataset". In order to have a balanced and complete test set, 50 ROIs were chosen among the BR and AWT classes, that surely contained airways, while the other 50 were selected from the HLT class. The performance of the model was scored as follows: correct if all the airways were correctly segmented, incorrect otherwise Out of the 100 ROIs selected, 58 contained at least one airway and 42 did not. The model correctly recognized the 42 ROIs without airways, and correctly segmented 51 out of the 58 ROIs with airways segmentation errors were present in 7 ROIs: in 4 cases the segmentation had more airways than the original image, while in the other 3 they were fewer or partially segmented. The pure model showed an overall accuracy of 93%.

In Figure 3.1 some examples of the model segmentation are reported.

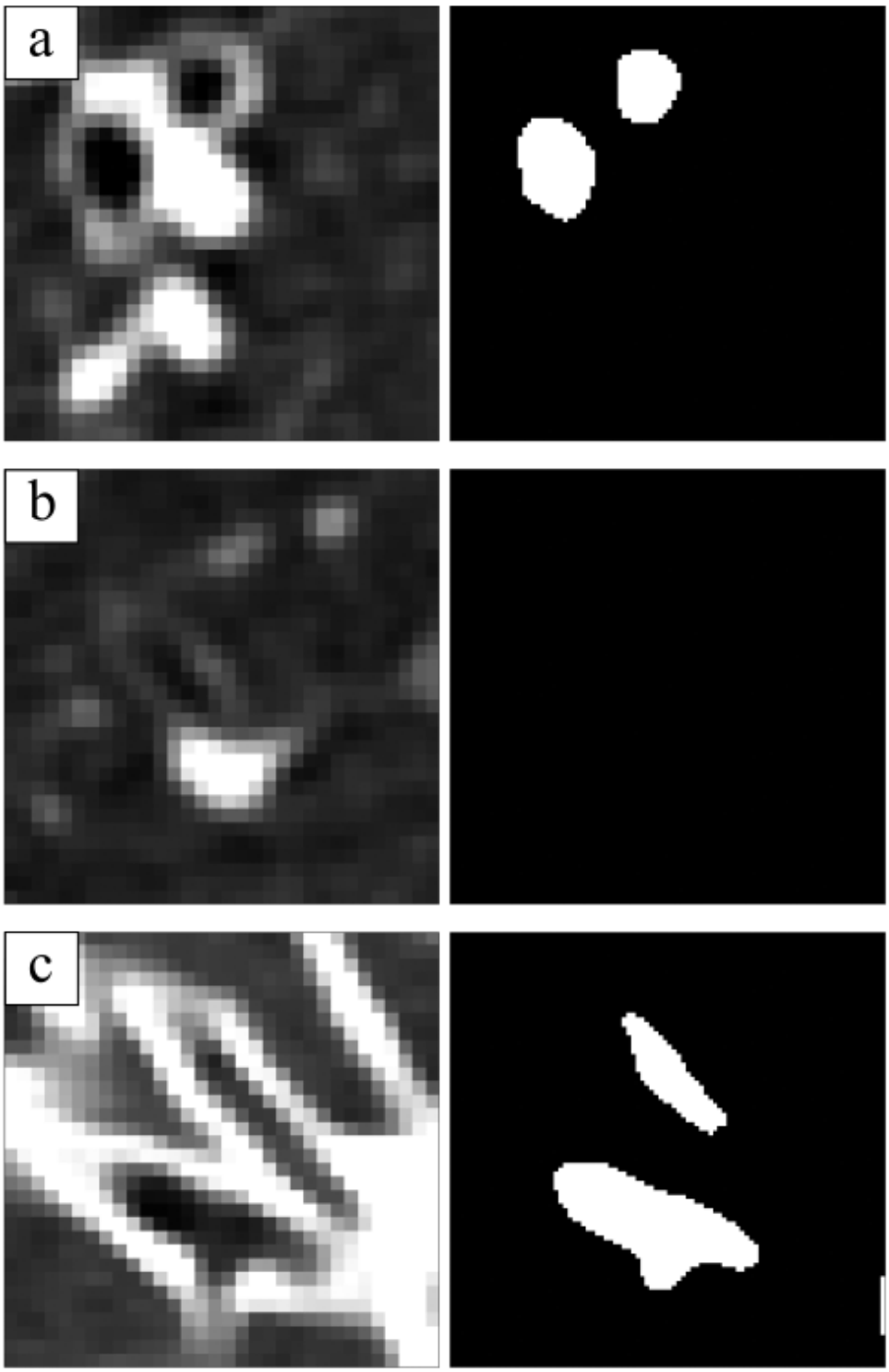


Figure 3.1: ***(a)*** The airways are correctly segmented. ***(b)*** In this case the model did not predict the presence of an airway in the original ROI. ***(c)*** Even if an airway is correctly segmented, the other one is just partially present in the model output, thus it was considered a mistake.

3.2. Classifiers Evaluation

The performance of the classifier models were evaluated as shown in Figure 3.2.

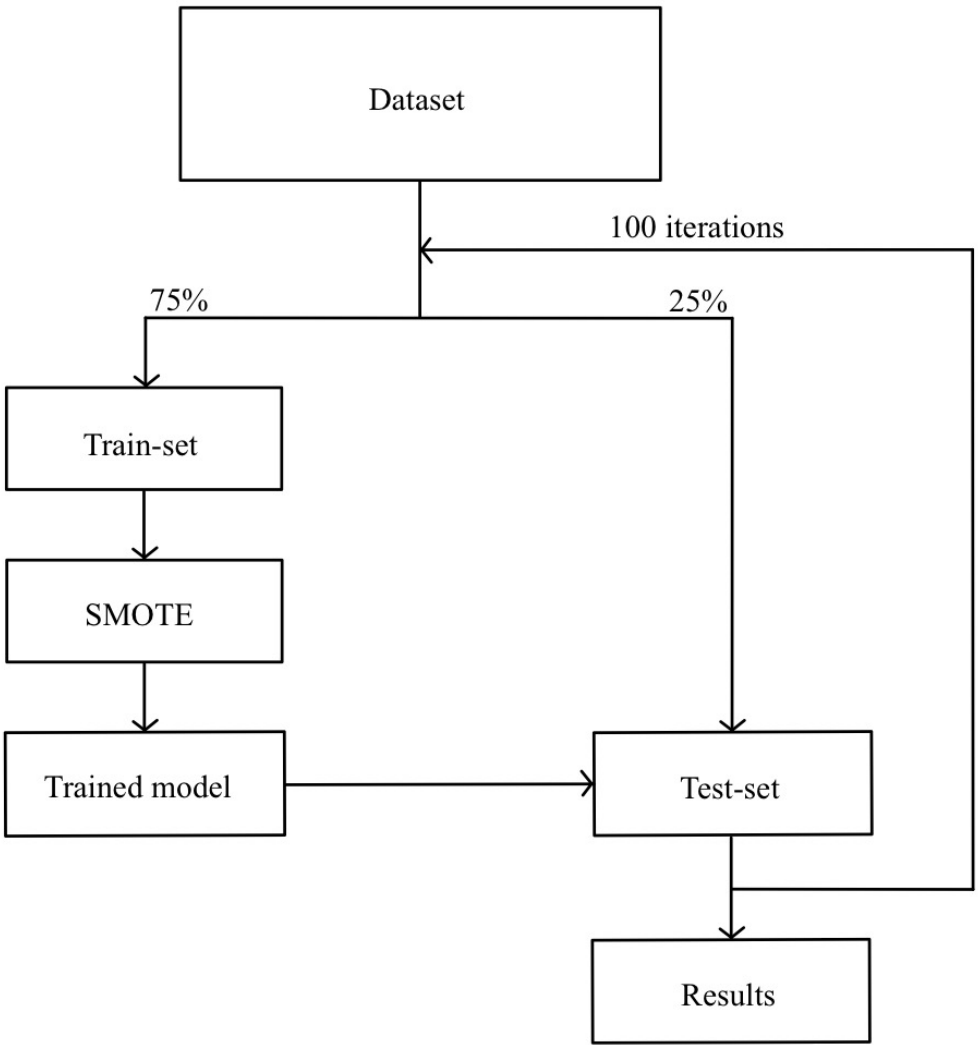


Figure 3.2: Procedure followed to evaluate the models performance.

The main dataset is first divided in two sets: 75% of the data are used for the train-set, while the remaining 25% for the test-set. As the train-set was imbalanced, SMOTE was applied to create synthetic samples in order to balance it. At the end of the splitting process the train-set was composed of 354 true observations, and after SMOTE of 496 samples equally spread among the 4 classes. Then, the model was trained and the performance was verified on the test-set. The performance of the model is strictly related

to how the samples are divided into the train and test-set, for this reason, the procedure described above was repeated 100 times to reduce the impact of the random splitting process. At the end of each iteration, recall, precision, specificity and f1 score were computed and thus the performance of the model are expressed as mean ± standard deviation.

In Figure 3.3 is shown an example of how precision, recall and specificity are computed in a multi-class problem.

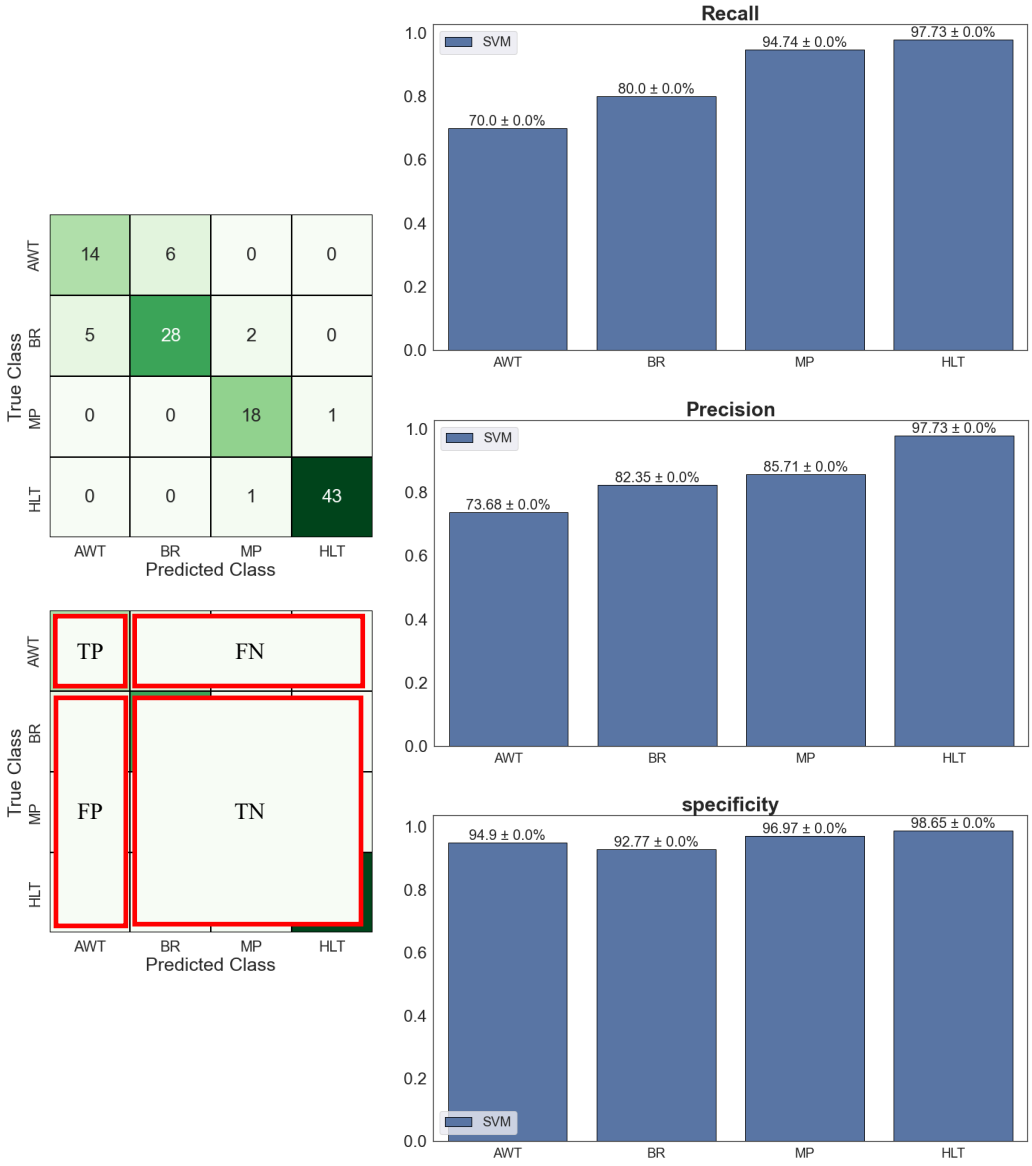


Figure 3.3: In a confusion matrix, the elements belonging to the same row are the true values, while the one belonging to the same columns are the elements predicted for that class. From the confusion matrix, it is possible to compute precision, recall and specificity for each class.

3.2.1. SVM vs Gaussian Naive Bayes

In Figure 3.4, the comparison of recall, precision and specificity between the two models is reported.

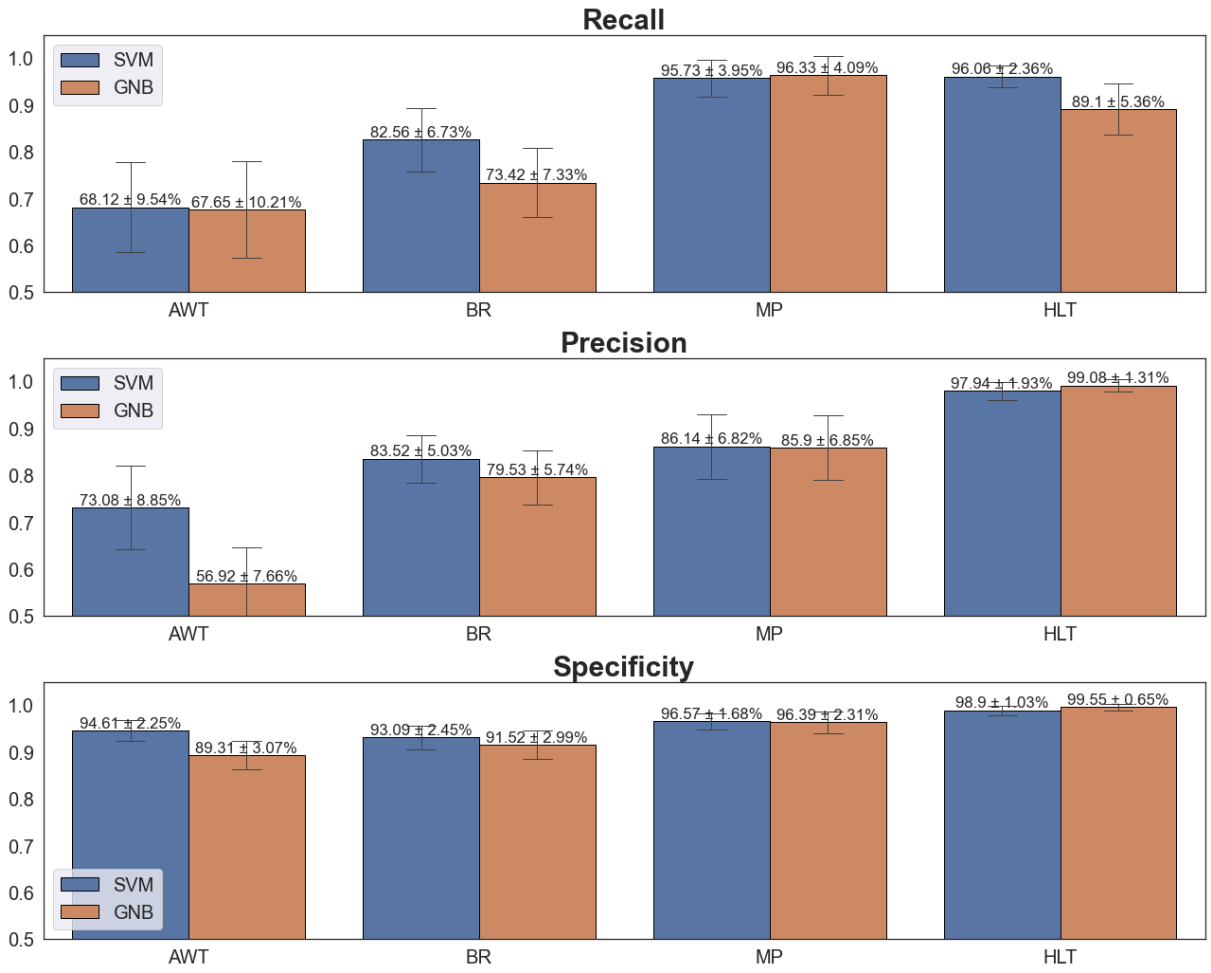


Figure 3.4: The three bar plots show the performance of the SVM and GNB. The metrics were recall, precision and specificity. The metrics were evaluated on 16x16 pixels regions. SVM had better recall, precision and specificity than the Bayesian model in almost every class.

The overall performance of the model was measured with the f1 score, which was equal to 82.32 ± 3.24% for the Gaussian Naive Bayes, and 87.06 ± 2.54% for the SVM.

The SVM approach had a better performance than the Bayesian one, and thus was chosen for the next steps. A summary of the Bayesian and SVM models are reported respectively in Table 3.1 and Table 3.2.

Bayesian	AWT	BR	MP	HLT
Recall	67.65±10.21%	73.42±7.33%	96.33±4.09%	89.10±5.36%
Precision	56.92±7.66%	79.53±5.74%	85.90±6.85%	99.08±1.31%
Specificity	89.31±3.07%	91.52±2.99%	96.39±2.31%	99.55±0.65%
f1: 82.32 ± 3.24%				

Table 3.1: Summary of Bayesian model performance.

SVM	AWT	BR	MP	HLT
Recall	68.12±9.54%	82.56±6.73%	95.73±3.95%	96.06±2.36%
Precision	73.08±8.85%	83.52±5.03%	86.14±6.82%	97.94±1.93%
Specificity	94.61±2.25%	93.09±2.45%	96.57±1.68%	98.90±1.03%
f1: 87.06 ± 2.54%				

Table 3.2: Summary of SVM model performance.

3.2.2. ROI Size Comparison

The performance of SVM was then tested on ROIs of different size: 16x16 pixels or 32x32 pixels. Smaller sized ROIs were discarded because they did not provide sufficient morphological information, due to their reduced field of view and also because airways usually don't fit inside such a small patch. The results are shown in Figure 3.5.

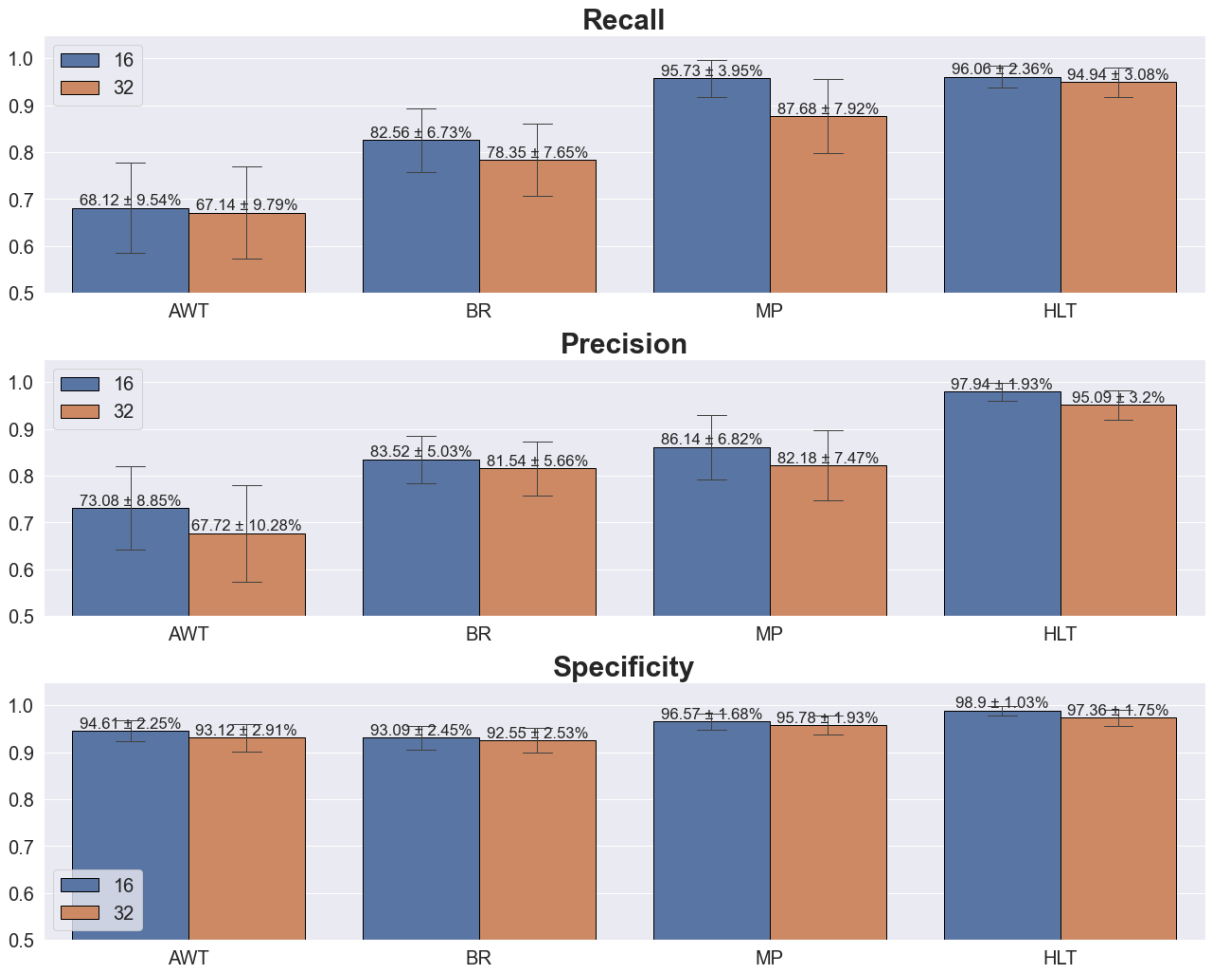


Figure 3.5: Recall, precision and specificity are computed with SVM on ROIs of different sizes, demonstrating that the 16x16 ROI has an overall better performance in all the classes compared to the 32x32.

The model trained and tested on ROI of size 16x16 pixels showed a better performance than the one developed on 32x32 ROI. This could be due to the fact that a 32x32 ROI covers more space and more than one class might appear in the same ROI, affecting the performance of the model. Hence, the model selected for the evaluation of the lung volume is a SVM with a ROI sized 16x16 pixels.

In Table 3.3 the comparison of the model trained and tested on the two different ROI size is summarized.

	ROI	AWT	BR	MP	HLT
Recall	16	68.12±9.54%	82.56±6.73%	95.73±3.95%	96.06±2.36%
	32	67.14±9.79%	78.35±7.65%	87.68±7.92%	94.94±3.08%
Precision	16	73.08±8.85%	83.52±5.03%	86.14±6.82%	97.94±1.93%
	32	67.72±10.28%	81.54±5.66%	82.18±7.47%	95.09±3.20%
Specificity	16	94.61±2.25%	93.09±2.45%	96.57±1.68%	98.90±1.03%
	32	93.12±2.91%	92.55±2.53%	95.78±1.93%	97.36±1.75%
f1 on 16x16 ROI: 87.06 ± 2.54%					
f1 on 32x32 ROI: 83.84 ± 2.99%					

Table 3.3: Comparison between the performance of the classifier on 16x16 and 32x32 ROI. The better performance of the 16x16 ROI is reflected by the higher fl score.

3.2.3. Impact of Airways Features

One of the aims of the thesis was to check if the use of information related to the airways could improve the performance of a texture analysis approach. To do so, the feature selection process described in Subsection 2.3.4 was applied on two features sets, the first one included the features related to the airways, while the latter did not. The first set of features, obviously, coincided with the final 8 features used for the evaluation of the models, while the other had only 5 features: Mean, Standard Deviation, Skewness, Contrast, GGO. The performance of the two models is shown in figure 3.6.

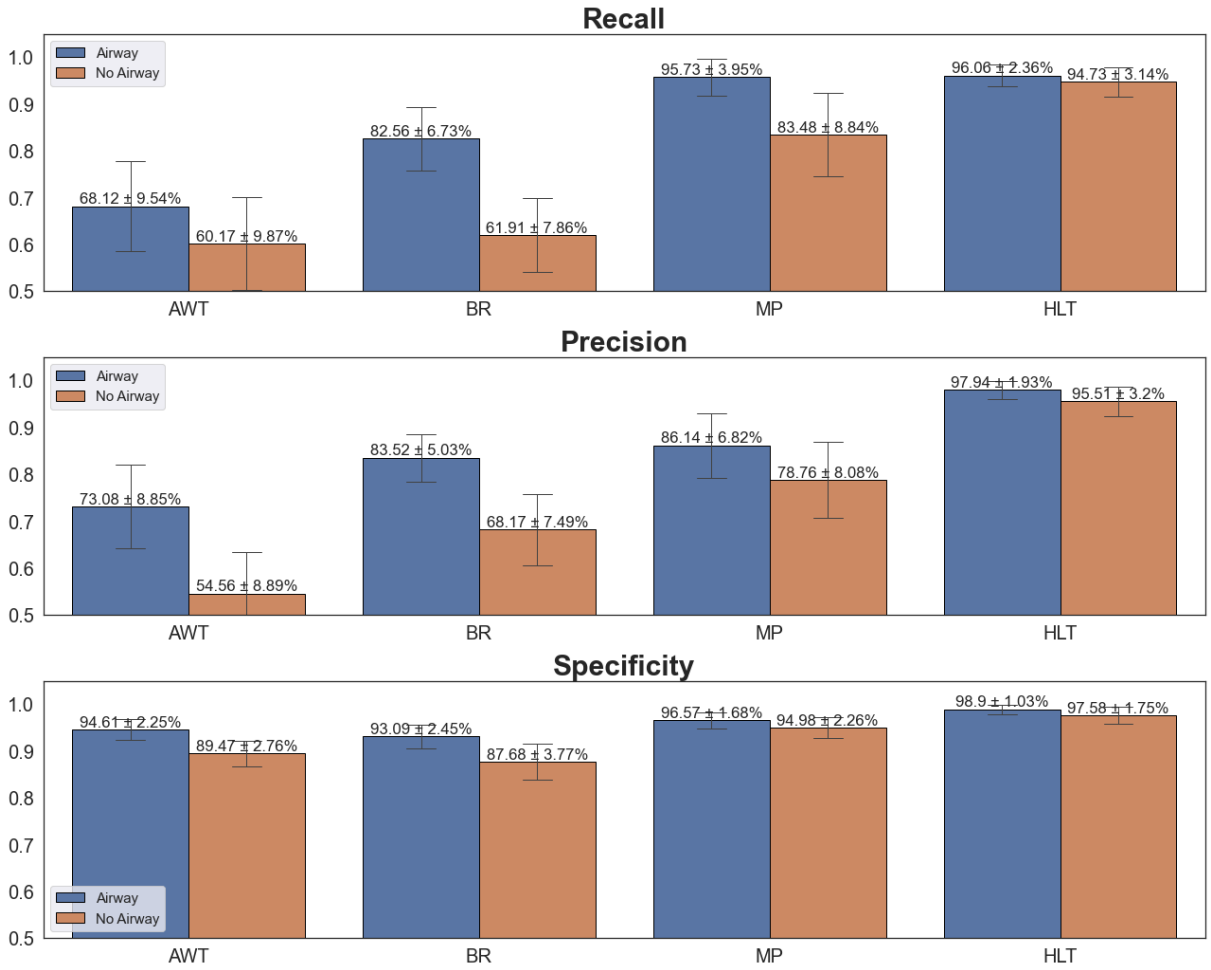


Figure 3.6: The features related to the airways had a strong impact on the correct classification of AWT, BR and even MP. The model trained on the set without airways features shows a worse performance in all the classes, for all the metrics.

As one could expect, the information related to the airways leaded to an increased performance of the model in the classification of AWT an BR. Surprisingly, the presence of these features had a positive impact also in the classification of MP.

In Table 3.4 the comparison between the two dataset is summarized.

	Airways	AWT	BR	MP	HLT
Recall	Yes	68.12±9.54%	82.56±6.73%	95.73±3.95%	96.06±2.36%
	No	60.17±9.87%	61.91±7.86%	83.48±8.84%	94.73±3.14%
Precision	Yes	73.08±8.85%	83.52±5.03%	86.14±6.82%	97.94±1.93%
	No	54.56±8.89%	68.17±7.49%	78.76±8.08%	95.51±3.20%
Specificity	Yes	94.61±2.25%	93.09±2.45%	96.57±1.68%	98.90±1.03%
	No	89.47±2.76%	87.68±3.77%	94.98±2.26%	97.58±1.75%
f1 with airways info : 87.06 ± 2.54%					
f1 without airways info : 76.93 ± 3.52%					

Table 3.4: Comparison between datasets with and without airways information.

3.2.4. Cross-Validation Results

Cross-validation was used in order to search for the best "C" parameter for the SVM, without having knowledge leaking from the test set into our model. The different values of C tried were: 1, 2, 5. The results are shown in Table 3.5.

	C	AWT	BR	MP	HLT
Recall	1	53.94±14.70%	88.70±4.79%	95.38±6.15%	95.17±2.99%
	2	53.94±14.70%	87.71±3.88%	93.85±8.97%	96.0±3.58%
	5	57.27±13.37%	86.75±5.62%	93.85±8.97%	96.0±3.58%
Precision	1	74.31±6.32%	79.79±8.46%	86.06±8.14%	99.23±1.54%
	2	72.70±9.78%	79.67±8.64%	87.34±9.63%	98.43±1.92%
	5	72.13±11.4%	81.13±7.95%	88.19±8.36%	98.43±1.92%
Specificity	1	96.27±1.27%	89.90±4.80%	96.20±2.75%	99.57±0.87%
	2	95.93±1.73%	89.90±4.80%	96.55±3.08%	99.12±1.08%
	5	95.25±2.92%	90.70±4.75%	96.89±2.53%	99.12±1.08%
f1 with C = 1: 85.90 ± 3.45%					
f1 with C = 2: 85.64 ± 4.33%					
f1 with C = 5: 86.07 ± 3.45%					

Table 3.5: Performance of SVM with different C values.

As there wasn't any significant increase in performance, C was set equal to 1.

3.3. Patients Scoring

The analysis was performed on the whole CT lung volume.

The lungs are then evaluated with our algorithm, the average time to elaborate an HRCT image is about 5 minutes. In Figure 3.7 two representative patients are shown. Out of the whole volume, three lung levels are reported.

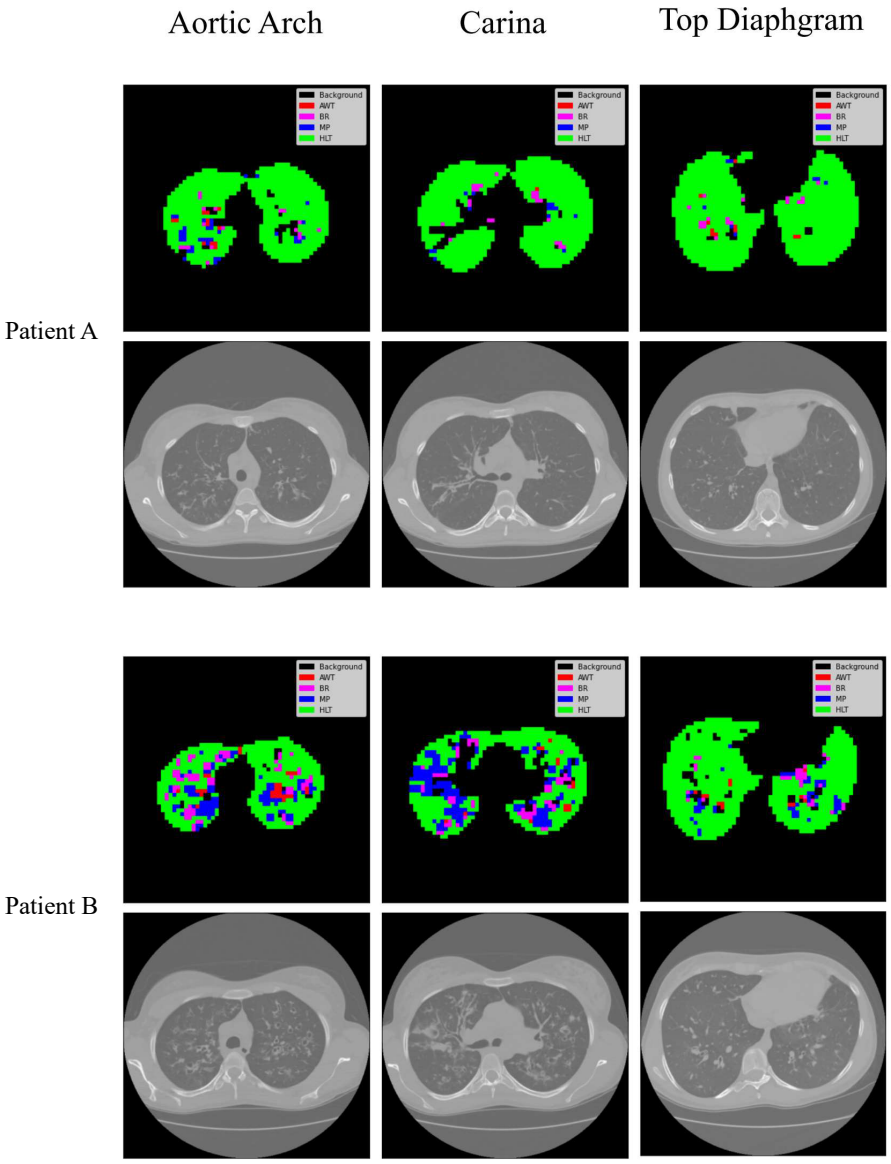


Figure 3.7: The images of the two representative patients are shown at three lung levels: aortic arch, carina and top diaphragm.

The results are provided as percentage of each class over the whole lung volume. (AWT%, BR%, MP%, HLT%) and the percentage of airways over the lung volume (vol_a%). The output of the algorithm for the two representative patients is summarized in Table 3.6 and Table 3.7.

Score	Presence
AWT%	0.15
BR%	0.89
MP%	0.44
HLT%	98.52
vol_a%	0.05

Table 3.6: Patient A lung composition.

Score	Presence
AWT%	0.28
BR%	2.50
MP%	4.18
HLT%	93.04
vol_a%	0.12

Table 3.7: Patient B lung composition.

3.4. Comparison with Radiological Scores

The quantitative data obtained with volume analysis were then compared with radiological scores in the group of patients with CF lung disease. The correlation between AWT%, BR%, MP% and radiological scores (AWT, BR, MP) was compared with Pearson correlation coefficient. The result of the correlation was reported in Table 3.8.

Scores	ρ AWT%	ρ BR%	ρ MP%
AWT	**0.72****	-	-
BR	-	**0.70****	-
MP	-	-	**0.55***

Table 3.8: Correlations results between radiological scores and the presence of patterns in HRCT. Significant values (*: p-value < 0.05; **: p-value < 0.01) are highlighted in bold.

The results of the lung volume analysis showed a good level of correlation with manual scores of the radiologist. Strong positive correlation was found between AWT score and AWT% ($\rho = 0.72$; p-value < 0.01) and BR score and BR% ($\rho = 0{,}70$; p-value < 0.01), while only moderate positive correlation was found for MP score and MP% ($\rho = 0.55$; p-value < 0.05).

Figure 3.8 shows the relation between BR, AWT and MP scores with BR%, AWT% and MP%.

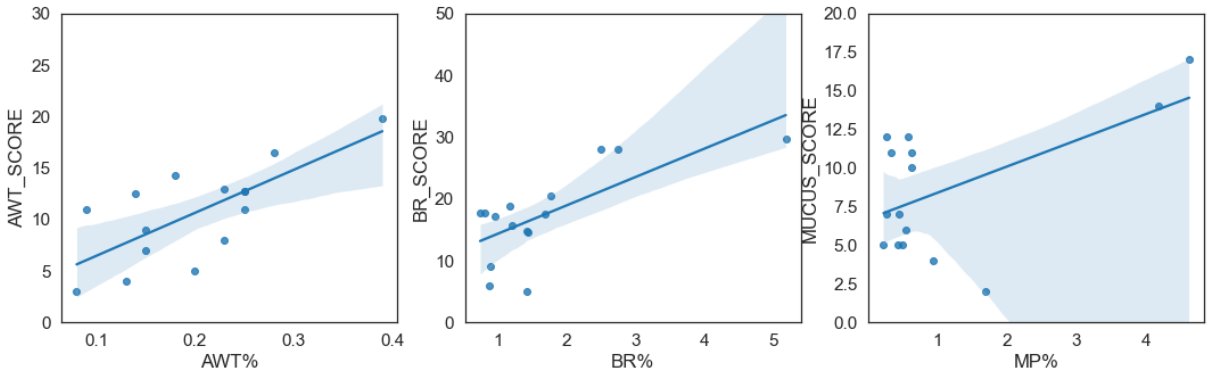


Figure 3.8: Comparison between BR, AWT, MP scores with BR%, AWT%, MP%. The translucent area around the regression model is the confidence interval at 95%.

3.5. Comparison with PFTs

The correlation between AWT%, BR%, MP%, HLT%, vol_a% and PFTs was compared with Spearman correlation coefficient in 15 patients. The results of the correlation are reported in Table 3.9 and significant results are highlighted in bold.

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***
$FEF_{25\%-75\%}$ (L)	-0,03	-0,36	-0,18	0,38	-0,15
%$FEF_{25\%-75\%}$	-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 3.9: Correlations results between PFTs and the percentage of each pattern in HRCT. Significant values (*: p-value < 0.05; **: p-value < 0.01) are highlighted in bold.

No significant correlation was found between AWT%, MP% and any of the PFTs. BR showed moderate correlation with %FEV_1 ($\rho = 0{,}53$; p-value < 0.05) and %FVC ($\rho = 0{,}60$; p-value < 0.05). HLT% presented moderate correlation with FEV_1 ($\rho = 0{,}58$; p-value < 0.05), %FEV_1 ($\rho = 0{,}52$; p-value < 0.05) and strong correlation with %FVC ($\rho = 0{,}68$; p-value < 0.01). Vol_a% showed significant correlation only with %FVC ($\rho = 0{,}52$; p-value < 0.05). In Table 3.9 are highlighted only the values which presented a p-value < 0.05. The correlation coefficient between LCI 2,5% norm and vol_a% had a p-value $= 0.08$.

In Figure 3.9, the scatterplots between %FEV_1, %FVC, %$FEF_{25\%-75\%}$, LCI 2,5% norm and AWT%, BR%, MP% and vol_a% are reported.

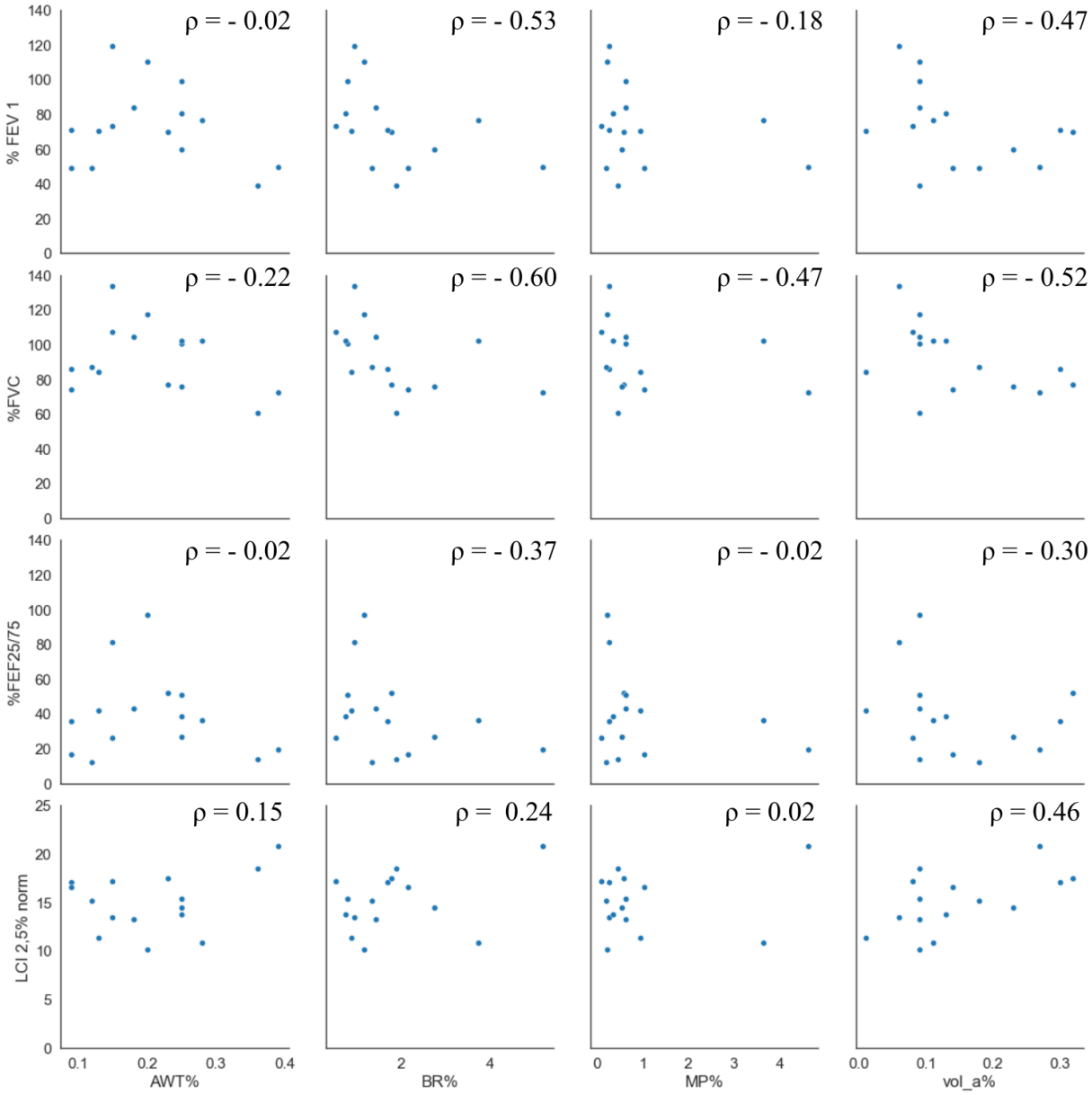


Figure 3.9: The scatterplots for each pair between $\%FEV_1$, %FVC, $\%FEF_{25\%-75\%}$, LCI 2,5% norm and AWT%, BR%, MP% and vol_a%.

4 | Conclusions and Future Developments

The scope of this work was to create an automated algorithm based on texture analysis and machine learning to evaluate pathological patterns in patient affected by cystic fibrosis lung disease. Other automated approaches found in literature exploit the use of convolutional neural network (CNN) to achieve the same scope. Marques et al. [74] proposed a model that was a cascade composition of two CNNs: the first one identified the presence of abnormal tissue and the latter the type of structural abnormalities. The first network could detect the presence of the disease with an accuracy of 0.94, while the cascade model presented an f1 score equal to 0.33, averaged over the classes.

In addition to common textural descriptors (thresholds, first order statistics, second order statistics), features derived from the airways were developed as surrogates of radiological indexes used in the evaluation of CT scans in patients affected by cystic fibrosis lung disease. This was done to improve the ability of the model to correctly identify and differentiate two of the four classes present in the dataset: BR and AWT. Two methods were implemented to segment the airways within the lungs. The first one was based on the identification of the contours inside of the ROIs, but it required the constant tweaking of parameters and hence it was not appropriate for an automated algorithm. The second method exploited the use of a UNET, that was trained on images of segmented airways. The performance of the UNET was tested on 100 random samples taken from our dataset, achieving an accuracy of 93%. Thus, the UNET output was used to propose airways candidates which still needed to be validated. An airway was accepted if it was a low-density area, completely surrounded by high attenuation structures.

The performance of two models, SVM and GNB, were tested for comparison. The SVM showed a better performance than GNB in all the metrics measured: recall, precision, specificity and f1 score. As SVM presented a significantly higher f1 score (SVM = 87.06 $\pm$ 2.54%, GNB = 82.32 $\pm$ 3.24%), it was selected for the following steps of the work.

Two ROI sizes (16x16 and 32x32) were tested. The model trained and tested 16x16 on

dataset presented an f1 score higher than the 32x32 (f1 = 87.06 ± 2.54%, 83.84 ± 2.99%, respectively for 16x16 and 32x32 ROIs).

The impact of the airways features was also investigated by comparing the performance of the SVM on two features dataset: the first one included airways features, while the latter did not. The model trained on the dataset with airways features showed the best recall, precision and specificity on the BR and AWT classes. Moreover, the MP class profited from airways features, as also its metrics improved. The presence of the airways features sensitively increased the performance of the model (f1 = 87.06 ± 2.54%, 76.93 ± 3.52%, respectively of the dataset with and without airways features).

The model was then used to evaluate the HRCT scans in a group of patients affected by cystic fibrosis lung disease. The volume of each class (AWT%, BR%, MP% and HLT%) as well as the volume of the airways (vol_a%) were computed as percentage of the total lung volume.

The dataset of HRCTs included 15 patients with CF lung disease. Pearson correlation coefficient was used to investigate the relationship between the volume of pathological classes automatically quantified with the developed algorithm (AWT%, BR% and MP%) and the visual radiological score (AWT, BR and MP). Strong positive correlation was found between AWT score and AWT% ($\rho = 0.72$; p-value < 0.01) and BR score and BR% ($\rho = 0.70$; p-value < 0.01), while only moderate positive correlation was found for MP score and MP% ($\rho = 0.55$; p-value < 0.05). These results suggest that our model is able to detect the presence of pathological pattern, but some discrepancies are present between the radiological scores and the pattern evaluation. For the radiological score, the radiologist assessed the severity and the extent of pathological patterns inside of the HRCT. Thus, a severe but localized pattern can present a higher, or similar, radiological score of a mild but diffused pattern. As our model evaluates only the percentage of a pattern inside of the lungs, is not able to consider whether it presents a severe or mild disease. In some cases the Frangi's filter confused MP for vessels and, as a consequence, MP might be underrepresented during our evaluation. Thus, other strategies for vessels segmentation should be explored. Moreover, the training dataset included a limited number of ROIs, thus we expect improvement with a larger dataset, balanced between the classes.

The correlation between our results and PFTs was also investigated in a group of 15 patients, to understand the relationship between structural and functional lung impairment. Negative correlation was found between BR% and PTFs (%FEV_1: ρ = -0.53, p-value $<$ 0.05; %FVC: ρ = -0.60, p-value < 0.05), whereas a positive correlation was found between HLT% and PFTs (FEV_1: $\rho = 0.58$, p-value < 0.05; %FEV_1: $\rho = 0.52$, p-value < 0.05;

%FVC: $\rho = 0.68$, p-value < 0.01). These results are reasonable as BR is a pathological pattern that causes a decrease in the PFTs of a subject, and a larger extent of healthy tissue (HLT%) may indicate a better lung function. The LCI showed a positive correlation (near significance, p-value $= 0.08$) with the percent volume of airways (vol_a%), suggesting that a larger volume of the airways tree (which is present in case of bronchiectasis) is an indicator of worse lung function.

This work presents some limitations. First, our dataset was composed of 472 ROIs. A bigger dataset might help in improving the performance of the model and limiting the impact of the SMOTE algorithm. Nevertheless, we notice that a bigger dataset doesn't directly imply a boost in the performance of the model because also the quality of the data is relevant. The identification of pathological patterns and the evaluation of radiological scores depends on the work of radiologists. Hence, a bigger dataset has more chance to include the effects of inter and intra-operator variability. When comparing the automatic disease quantification with PFTs, we notice that the patients analyzed in this thesis were enrolled in a clinical trial. Thus, our algorithm was tested only on patients corresponding to the enrolment criteria of the trial. In particular, only patients with mild or moderate disease were included, while end-stage subjects were not present.

Further work could be oriented in improving the performance of our model. Our approach is able to completely explore the whole lung volume, but it is not a 3D analysis, as each slice in the HRCT is treated singularly during the analysis. Not only the standard textural features could be adapted to a three dimensional approach, but also the airways features might gain benefit from it. The proposed UNET works only on 2D images without exploiting the continuity of the airways in the segmentation process. Moreover, a 3D UNET could detect with higher precision not only the airways lumen, but it could also be trained to identify the airways wall and the corresponding vessels, to be used to evaluate the airway to artery ratio, the thickness of the airway wall with respect to the lumen and the presence of airway tapering. These parameters are usually used by the radiologist in the correct identification of AWT and BR. When applying our model to CT scans from patients, a further step would be to apply the algorithm to longitudinal data in patients under treatment, to evaluate the additional value of this approach in a clinical trial.

Bibliography

[1] Blanka Klimova, Kamil Kuca, Michal Novotny, and Petra Maresova. Cystic fibrosis revisited – a review study. *Medicinal Chemistry*, 13(2):102–109, January 2017.

[2] Kris De Boeck. Cystic fibrosis in the year 2020: A disease with a new face. *Acta Paediatrica*, 109(5):893–899, January 2020.

[3] Donald R. VanDevanter, Jennifer S. Kahle, Amy K. O'Sullivan, Slaven Sikirica, and Paul S. Hodgkins. Cystic fibrosis in young children: A review of disease manifestation, progression, and response to early treatment. *Journal of Cystic Fibrosis*, 15(2):147–157, March 2016.

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

[5] Frank J. Accurso. Early pulmonary disease in cystic fibrosis. *Current Opinion in Pulmonary Medicine*, 3(6):400–403, November 1997.

[6] Wieying Kuo, , Adria Perez-Rovira, Harm Tiddens, and Marleen de Bruijne. Airway tapering: an objective image biomarker for bronchiectasis. *European Radiology*, 30(5):2703–2711, February 2020.

[7] Valentina Ferraro, Eleni-Rosalina Andrinopoulou, Anna Marthe Margaretha Sijbring, Eric G. Haarman, Harm A.W.M. Tiddens, and Marielle W.H. Pijnenburg. Airway–artery quantitative assessment on chest computed tomography in paediatric primary ciliary dyskinesia. *ERJ Open Research*, 6(3):00210–2019, July 2020.

[8] Fenker Daniel E, McDaniel Cameron T, Panmanee Warunya, Panos Ralph J, Sorscher Eric J, Sabusap Carleen, Clancy John P, and Hassett Daniel J. A comparison between two pathophysiologically different yet microbiologically similar lung diseases: Cystic fibrosis and chronic obstructive pulmonary disease. *International Journal of Respiratory and Pulmonary Medicine*, 5(2), December 2018.

[9] J. Wanger, J. L. Clausen, A. Coates, O. F. Pedersen, V. Brusasco, F. Burgos, R. Casaburi, R. Crapo, P. Enright, C. P. M. van der Grinten, P. Gustafsson, J. Han-

kinson, R. Jensen, D. Johnson, N. MacIntyre, R. McKay, M. R. Miller, D. Navajas, R. Pellegrino, and G. Viegi. Standardisation of the measurement of lung volumes. *European Respiratory Journal*, 26(3):511–522, September 2005.

[10] Theodore G. Liou and Richard E. Kanner. Spirometry. *Clinical Reviews in Allergy & Immunology*, 37(3):137–152, April 2009.

[11] Jeffrey M Haynes, David A Kaminsky, Sanja Stanojevic, and Gregg L Ruppel. Pulmonary function reference equations: A brief history to explain all the confusion. *Respiratory Care*, 65(7):1030–1038, March 2020.

[12] Rhonda Szczesniak, Sonya L. Heltshe, Sanja Stanojevic, and Nicole Mayer-Hamblett. Use of FEV1 in cystic fibrosis epidemiologic studies and clinical trials: A statistical perspective for the clinical researcher. *Journal of Cystic Fibrosis*, 16(3):318–326, May 2017.

[13] H. A. Kerstjens, B. Rijcken, J. P. Schouten, and D. S. Postma. Decline of FEV1 by age and smoking status: facts, figures, and fallacies. *Thorax*, 52(9):820–827, September 1997.

[14] E. Kerem, L. Viviani, A. Zolin, S. MacNeill, E. Hatziagorou, H. Ellemunter, P. Drevinek, V. Gulmans, U. Krivec, and H. Olesen and. Factors associated with FEV1 decline in cystic fibrosis: analysis of the ECFS patient registry. *European Respiratory Journal*, 43(1):125–133, April 2013.

[15] Kathryn A. Ramsey, Tim Rosenow, Lidija Turkovic, Billy Skoric, Georgia Banton, Anne-Marie Adams, Shannon J. Simpson, Conor Murray, Sarath C. Ranganathan, Stephen M. Stick, and Graham L. Hall. Lung clearance index and structural lung disease on computed tomography in early cystic fibrosis. *American Journal of Respiratory and Critical Care Medicine*, 193(1):60–67, January 2016.

[16] Magdalena Kołodziej, Michael J. de Veer, Marian Cholewa, Gary F. Egan, and Bruce R. Thompson. Lung function imaging methods in cystic fibrosis pulmonary disease. *Respiratory Research*, 18(1), May 2017.

[17] E. Marije Bakker, Gerard J.J.M. Borsboom, Els C. van der Wiel-Kooij, Daan Caudri, Margaret Rosenfeld, and Harm A.W.M. Tiddens. Small airway involvement in cystic fibrosis lung disease: Routine spirometry as an early and sensitive marker. *Pediatric Pulmonology*, 48(11):1081–1088, February 2013.

[18] Nowaf Y Alobaidi, Mohammed Almeshari, James Stockley, Robert Andrew Stockley, and Elizabeth Sapey. Small airway function measured using forced expiratory flow

between 25% and 75% of vital capacity and its relationship to airflow limitation in symptomatic ever-smokers: a cross-sectional study. *BMJ Open Respiratory Research*, 9(1):e001385, October 2022.

[19] Harm A.W.M. Tiddens, Michael Puderbach, Jose G. Venegas, Felix Ratjen, Scott H. Donaldson, Stephanie D. Davis, Steven M. Rowe, Scott D. Sagel, Mark Higgins, and David A. Waltz. Novel outcome measures for clinical trials in cystic fibrosis. *Pediatric Pulmonology*, 50(3):302–315, December 2014.

[20] Gloria Tridello, Sonia Volpi, Baroukh M. Assael, Ilaria Meneghelli, Marianna Passiu, and Maria Circelli. Lung function comparison between two decades in cystic fibrosis children: A single centre study. *Pediatric Pulmonology*, 50(12):1237–1243, September 2015.

[21] Padmaja Subbarao, Carlos Milla, Paul Aurora, Jane C. Davies, Stephanie D. Davis, Graham L. Hall, Sonya Heltshe, Philipp Latzin, Anders Lindblad, Jessica E. Pittman, Paul D. Robinson, Margaret Rosenfeld, Florian Singer, Tim D. Starner, Felix Ratjen, and Wayne Morgan. Multiple-breath washout as a lung function test in cystic fibrosis. a cystic fibrosis foundation workshop report. *Annals of the American Thoracic Society*, 12(6):932–939, June 2015.

[22] S. Lum, P. Gustafsson, H. Ljungberg, G. Hulskamp, A. Bush, S. B Carr, R. Castle, A. f. Hoo, J. Price, S. Ranganathan, J. Stroobant, A. Wade, C. Wallis, H. Wyatt, and J. Stocks and. Early detection of cystic fibrosis lung disease: multiple-breath washout versus raised volume tests. *Thorax*, 62(4):341–347, January 2007.

[23] BreAnna Kinghorn, Sharon McNamara, Alan Genatossio, Erin Sullivan, Molly Siegel, Irma Bauer, Charles Clem, Robin C. Johnson, Miriam Davis, Anne Griffiths, William Wheeler, Katherine Johnson, Stephanie D. Davis, Margaret W. Leigh, Margaret Rosenfeld, and Jessica Pittman. Comparison of multiple breath washout and spirometry in children with primary ciliary dyskinesia and cystic fibrosis and healthy controls. *Annals of the American Thoracic Society*, 17(9):1085–1093, September 2020.

[24] 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.

[25] Harm A. W. M. Tiddens and Tim Rosenow. What did we learn from two decades of chest computed tomography in cystic fibrosis? *Pediatric Radiology*, 44(12):1490–1495, August 2014.

[26] Anna Rybacka and Katrzyna Karmelita-Katulska. The role of computed tomography

in monitoring patients with cystic fibrosis. *Polish Journal of Radiology*, 81:141–145, April 2016.

[27] Rhonda Szczesniak, Lidija Turkovic, Eleni-Rosalina Andrinopoulou, and Harm A.W.M. Tiddens. Chest imaging in cystic fibrosis studies: What counts, and can be counted? *Journal of Cystic Fibrosis*, 16(2):175–185, March 2017.

[28] Frederick R Long, Roger S Williams, and Robert G Castile. Structural airway abnormalities in infants and young children with cystic fibrosis. *The Journal of Pediatrics*, 144(2):154–161, February 2004.

[29] Pim A. de Jong, Mark D. Ottink, Simon G. F. Robben, Maarten H. Lequin, Wim C. J. Hop, Johan J. E. Hendriks, Peter D. Paré, and Harm A. W. M. Tiddens. Pulmonary disease assessment in cystic fibrosis: Comparison of CT scoring systems and value of bronchial and arterial dimension measurements. *Radiology*, 231(2):434–439, May 2004.

[30] Wieying Kuo, Marleen de Bruijne, Jens Petersen, Kazem Nasserinejad, Hadiye Ozturk, Yong Chen, Adria Perez-Rovira, and Harm A. W. M. Tiddens. Diagnosis of bronchiectasis and airway wall thickening in children with cystic fibrosis: Objective airway-artery quantification. *European Radiology*, 27(11):4680–4689, May 2017.

[31] Harm A.W.M. Tiddens, Jennifer J. Meerburg, Menno M. van der Eerden, and Pierluigi Ciet. The radiological diagnosis of bronchiectasis: what's in a name? *European Respiratory Review*, 29(156):190120, June 2020.

[32] Jürgen Schäfer, Matthias Griese, Ravishankar Chandrasekaran, Sanjay H. Chotirmall, and Dominik Hartl. Pathogenesis, imaging and clinical characteristics of CF and non-CF bronchiectasis. *BMC Pulmonary Medicine*, 18(1), May 2018.

[33] Carlos H. Lugo-Olivieri, Philippe A. Soyer, and Elliot K. Fishman. Cystic fibrosis. *Clinical Imaging*, 22(5):346–354, September 1998.

[34] Jean-Paul Charbonnier, Esther Pompe, Camille Moore, Stephen Humphries, Bram van Ginneken, Barry Make, Elizabeth Regan, James D. Crapo, Eva M. van Rikxoort, and David A. Lynch. Airway wall thickening on CT: Relation to smoking status and severity of COPD. *Respiratory Medicine*, 146:36–41, January 2019.

[35] Victor Kim, , Wojciech R. Dolliver, Hrudaya P. Nath, Scott A. Grumley, Nina Terry, Asmaa Ahmed, Andrew Yen, Kathleen Jacobs, Seth Kligerman, and Alejandro A. Diaz. Mucus plugging on computed tomography and chronic bronchitis in chronic obstructive pulmonary disease. *Respiratory Research*, 22(1), April 2021.

[36] 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.

[37] 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.

[38] 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.

[39] Nynke R. Bouma, Hettie M. Janssens, Eleni-Rosalina Andrinopoulou, and Harm A. W. M. Tiddens. Airway disease on chest computed tomography of preschool children with cystic fibrosis is associated with school-age bronchiectasis. *Pediatric Pulmonology*, 55(1):141–148, September 2019.

[40] Don B. Sanders, Zhanhai Li, and Alan S. Brody. Chest computed tomography predicts the frequency of pulmonary exacerbations in children with cystic fibrosis. *Annals of the American Thoracic Society*, 12(1):64–69, January 2015.

[41] Martine Loeve, Wim C. J. Hop, Marleen de Bruijne, Peter T. W. van Hal, Phil Robinson, Moira L. Aitken, Jonathan D. Dodd, and Harm A. W. M. Tiddens. Chest computed tomography scores are predictive of survival in patients with cystic fibrosis awaiting lung transplantation. *American Journal of Respiratory and Critical Care Medicine*, 185(10):1096–1103, May 2012.

[42] A.K. Jain, P.W. Duin, and Jianchang Mao. Statistical pattern recognition: a review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(1):4–37, 2000.

[43] Alicia Chen, Ronald A. Karwoski, David S. Gierada, Brian J. Bartholmai, and Chi Wan Koo. Quantitative CT analysis of diffuse lung disease. *RadioGraphics*, 40(1):28–43, January 2020.

[44] Sundaresh Ram, Benjamin A. Hoff, Alexander J. Bell, Stefanie Galban, Aleksa B. Fortuna, Oliver Weinheimer, Mark O. Wielpütz, Terry E. Robinson, Beverley Newman, Dharshan Vummidi, Aamer Chughtai, Ella A. Kazerooni, Timothy D. Johnson, MeiLan K. Han, Charles R. Hatt, and Craig J. Galban. Improved detection of air

trapping on expiratory computed tomography using deep learning. *PLOS ONE*, 16(3):e0248902, March 2021.

[45] Olga Solyanik, Patrick Hollmann, Sabine Dettmer, Till Kaireit, Cornelia Schaefer-Prokop, Frank Wacker, Jens Vogel-Claussen, and Hoen oh Shin. Quantification of pathologic air trapping in lung transplant patients using CT density mapping: Comparison with other CT air trapping measures. *PLOS ONE*, 10(10):e0139102, October 2015.

[46] Robert M. Haralick, K. Shanmugam, and Its'Hak Dinstein. Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-3(6):610–621, November 1973.

[47] Robert J. Gillies, Paul E. Kinahan, and Hedvig Hricak. Radiomics: Images are more than pictures, they are data. *Radiology*, 278(2):563–577, February 2016.

[48] Elisa Scalco and Giovanna Rizzo. Texture analysis of medical images for radiotherapy applications. *The British Journal of Radiology*, 90(1070):20160642, February 2017.

[49] Adrien Depeursinge, Antonio Foncubierta-Rodriguez, Dimitri Van De Ville, and Henning Müller. Three-dimensional solid texture analysis in biomedical imaging: Review and opportunities. *Medical Image Analysis*, 18(1):176–196, January 2014.

[50] Meghan G. Lubner, Andrew D. Smith, Kumar Sandrasegaran, Dushyant V. Sahani, and Perry J. Pickhardt. CT texture analysis: Definitions, applications, biologic correlates, and challenges. *RadioGraphics*, 37(5):1483–1503, September 2017.

[51] William Rogers, Sithin Thulasi Seetha, Turkey A. G. Refaee, Relinde I. Y. Lieverse, Renée W. Y. Granzier, Abdalla Ibrahim, Simon A. Keek, Sebastian Sanduleanu, Sergey P. Primakov, Manon P. L. Beuque, Damiënne Marcus, Alexander M. A. van der Wiel, Fadila Zerka, Cary J. G. Oberije, Janita E van Timmeren, Henry C. Woodruff, and Philippe Lambin. Radiomics: from qualitative to quantitative imaging. *The British Journal of Radiology*, 93(1108):20190948, April 2020.

[52] K. Buch, B. Li, M.M. Qureshi, H. Kuno, S.W. Anderson, and O. Sakai. Quantitative assessment of variation in CT parameters on texture features: Pilot study using a nonanatomic phantom. *American Journal of Neuroradiology*, 38(5):981–985, March 2017.

[53] Youngmin Yoon, Taesung Hwang, Hojung Choi, and Heechun Lee. Classification of radiographic lung pattern based on texture analysis and machine learning. *Journal of Veterinary Science*, 20(4), 2019.

[54] Jinju Sun, Kaijun Liu, Haipeng Tong, Huan Liu, Xiaoguang Li, Yi Luo, Yang Li, Yun Yao, Rongbing Jin, Jingqin Fang, and Xiao Chen. CT texture analysis for differentiating bronchiolar adenoma, adenocarcinoma in situ, and minimally invasive adenocarcinoma of the lung. *Frontiers in Oncology*, 11, April 2021.

[55] Kiminobu Tanizawa, Tomohiro Handa, Sonoko Nagai, Toyohiro Hirai, Takeshi Kubo, Tsuyoshi Oguma, Isao Ito, Yutaka Ito, Kizuku Watanabe, Kensaku Aihara, Kohei Ikezoe, Toru Oga, Kazuo Chin, Takateru Izumi, and Michiaki Mishima. Clinical impact of high-attenuation and cystic areas on computed tomography in fibrotic idiopathic interstitial pneumonias. *BMC Pulmonary Medicine*, 15(1), July 2015.

[56] Chia-Hung Chen, Chih-Kun Chang, Chih-Yen Tu, Wei-Chih Liao, Bing-Ru Wu, Kuei-Ting Chou, Yu-Rou Chiou, Shih-Neng Yang, Geoffrey Zhang, and Tzung-Chi Huang. Radiomic features analysis in computed tomography images of lung nodule classification. *PLOS ONE*, 13(2):e0192002, February 2018.

[57] Guillaume Chassagnon, Clémence Martin, Pierre-Régis Burgel, Dominique Hubert, Isabelle Fajac, Nikos Paragios, Evangelia I. Zacharaki, Paul Legmann, Joel Coste, and Marie-Pierre Revel. An automated computed tomography score for the cystic fibrosis lung. *European Radiology*, 28(12):5111–5120, June 2018.

[58] Emily M. DeBoer, Waldemar Swiercz, Sonya L. Heltshe, Margaret M. Anthony, Paul Szefler, Rebecca Klein, John Strain, Alan S. Brody, and Scott D. Sagel. Automated CT scan scores of bronchiectasis and air trapping in cystic fibrosis. *Chest*, 145(3):593–603, March 2014.

[59] Guillaume Chassagnon, Evangelia I. Zacharaki, Sébastien Bommart, Pierre-Régis Burgel, Raphael Chiron, Séverine Dangeard, Nikos Paragios, Clémence Martin, and Marie-Pierre Revel. Quantification of cystic fibrosis lung disease with radiomics-based CT scores. *Radiology: Cardiothoracic Imaging*, 2(6):e200022, December 2020.

[60] Marcel Koenigkam Santos, Danilo Lemos Cruvinel, Marcelo Bezerra de Menezes, Sara Reis Teixeira, Elcio de Oliveira Vianna, Jorge Elias Júnior, and José Antonio Baddini Martinez. Quantitative computed tomography analysis of the airways in patients with cystic fibrosis using automated software: correlation with spirometry in the evaluation of severity. *Radiologia Brasileira*, 49(6):351–357, October 2016.

[61] Bishop Christopher M. *Pattern Recognition and Machine Learning*. Information Science and Statistics. Springer, New York, NY, 1 edition, August 2006.

[62] Roberto Grassi, Maria Paola Belfiore, Alessandro Montanelli, Gianluigi Patelli, Fabrizio Urraro, Giuliana Giacobbe, Roberta Fusco, Vincenza Granata, Antonella

Petrillo, Palmino Sacco, Maria Antonietta Mazzei, Beatrice Feragalli, Alfonso Reginelli, and Salvatore Cappabianca. COVID-19 pneumonia: computer-aided quantification of healthy lung parenchyma, emphysema, ground glass and consolidation on chest computed tomography (CT). *La radiologia medica*, 126(4):553–560, November 2020.

[63] Nikolas Adaloglou. `https://theaisummer.com/medical-image-python/`.

[64] William E. Lorensen and Harvey E. Cline. Marching cubes: A high resolution 3d surface construction algorithm. *ACM SIGGRAPH Computer Graphics*, 21(4):163–169, August 1987.

[65] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation, 2015.

[66] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[67] Shigang Liu, Jun Zhang, Yang Xiang, Wanlei Zhou, and Dongxi Xiang. A study of data pre-processing techniques for imbalanced biomedical data classification. *International Journal of Bioinformatics Research and Applications*, 16(3):290, 2020.

[68] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer. SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16:321–357, June 2002.

[69] Hui Han, Wen-Yuan Wang, and Bing-Huan Mao. Borderline-SMOTE: A new over-sampling method in imbalanced data sets learning. In *Lecture Notes in Computer Science*, pages 878–887. Springer Berlin Heidelberg, 2005.

[70] Haibo He, Yang Bai, Edwardo A. Garcia, and Shutao Li. ADASYN: Adaptive synthetic sampling approach for imbalanced learning. In *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*. IEEE, June 2008.

[71] Alejandro F. Frangi, Wiro J. Niessen, Koen L. Vincken, and Max A. Viergever. Multiscale vessel enhancement filtering. In *Medical Image Computing and Computer-Assisted Intervention — MICCAI'98*, pages 130–137. Springer Berlin Heidelberg, 1998.

[72] M. R. Miller. Standardisation of spirometry. *European Respiratory Journal*, 26(2):319–338, August 2005.

[73] Renee Jensen, Kent Green, Per Gustafsson, et al. *Standard Operating Procedure: Multiple Breath Nitrogen Washout*. Eco Medics AG, 2013.

[74] Filipe Marques, Marleen de Bruijne, Florian Dubost, Harm A. W. Tiddens, and Mariette Kemner van de Corput. Quantification of lung abnormalities in cystic fibrosis using deep networks. In Elsa D. Angelini and Bennett A. Landman, editors, *Medical Imaging 2018: Image Processing*. SPIE, March 2018.

List of Figures

List of Tables

Acknowledgements

Thanks to Prof. Andrea Aliverti for giving me the possibility to develop this master thesis.

A special thanks to Francesca Pennati for her support and advices in the realization of this work.