



**INSTITUTO POLITÉCNICO DE LISBOA**



ESCOLA SUPERIOR DE  
TECNOLOGIA DA SAÚDE  
DE LISBOA  
INSTITUTO POLITÉCNICO DE LISBOA

Instituto Superior de Engenharia de Lisboa  
Escola Superior de Tecnologia da Saúde de Lisboa

# **Previsão de volumes pulmonares não mobilizáveis com base em parâmetros espirométricos**

MARCO ANTÓNIO PACHECO PEREIRA

Trabalho Final de Mestrado para obtenção do grau de  
Mestre em Engenharia Biomédica

## Orientadores

Doutor Ricardo Teresa Ribeiro (ESTeSL – IPL)

Mestre Herminia Brites Dias (ESTeSL – IPL)

**dezembro de 2022**



INSTITUTO POLITÉCNICO DE LISBOA



ESCOLA SUPERIOR DE  
TECNOLOGIA DA SAÚDE  
DE LISBOA  
INSTITUTO POLITÉCNICO DE LISBOA

Instituto Superior de Engenharia de Lisboa  
Escola Superior de Tecnologia da Saúde de Lisboa

# Previsão de volumes pulmonares não mobilizáveis com base em parâmetros espirométricos

MARCO ANTÓNIO PACHECO PEREIRA

Trabalho Final de Mestrado para obtenção do grau de  
Mestre em Engenharia Biomédica

## Orientadores

Doutor Ricardo Teresa Ribeiro (ESTeSL – IPL)

Mestre Herminia Brites Dias (ESTeSL – IPL)

## Júri:

Presidente:

Doutor António Jorge Duarte de Castro Silvestre (ISEL)

Vogais:

Doutor André Valério Raposo Carreiro (Fraunhofer, Portugal – AICOS)

Doutor Ricardo Teresa Ribeiro (ESTeSL – IPL)

dezembro de 2022

## Agradecimentos

Ao Professor Ricardo Ribeiro, por me ter introduzido ao mundo da Ciência de Dados, através do seu exemplo, tanto como pessoa como Engenheiro Biomédico. Obrigado por ter tornado esta dissertação possível.

À Professora Hermínia Dias, pela paixão que tem pelas áreas que ensina e pela motivação que incute aos outros, levando a que façam mais e melhor. Obrigado pela disponibilidade e pelos ensinamentos que me concedeu ao longo destes últimos dez anos.

Ao serviço de Pneumologia do Hospital da Luz de Lisboa, mais concretamente à equipa de Técnicos de Cardiopneumologia, por me terem feito crescer pessoalmente e profissionalmente.

À minha mãe, por todo o amor e carinho que sempre me deste e que desejo retribuir. Espero tê-lo feito.

Ao meu pai, pelos valores que me ensinaste e que espero vir a ensinar aos meus filhos.

À minha irmã, por seres aquela pessoa com quem eu sei que poderei sempre contar.

À minha mulher e amor da minha vida, por seres mais do que a minha metade, por seres o meu todo.

Ao meu filho Lourenço, obrigado por fazeres com que tudo valha a pena.

*There is a theory which states that if anyone discovers exactly what the Universe is for and why it is here, it will instantly disappear and be replaced by something even more bizarre and inexplicable...*

*There is another theory which states that this has already hapenned.*

Douglas Adams em *The Hitchhiker's Guide to the Galaxy*

## Resumo

Definição do problema: A prova de função respiratória (PFR) mais acessível e utilizada atualmente é a espirometria. Contudo, é incapaz de medir os volumes pulmonares estáticos, essenciais para o diagnóstico de hiperinsuflação e de restrição pulmonar. As técnicas de aprendizagem automática (ML) têm ganho notoriedade no campo da medicina em grande parte graças à sua capacidade de prever ou classificar a partir de um grande número de exemplos já conhecidos.

Objetivo: Aplicar modelos supervisionados de regressão para os volumes pulmonares não mobilizáveis e de classificação automática das suas alterações, através de parâmetros biológicos, antropométricos e espirométricos, com avaliação do seu desempenho.

Metodologia: Estudo retrospectivo com 8140 PFR anonimizadas contendo dados biológicos, antropométricos, espirométricos e de pletismografia corporal total (PCT). Oito modelos foram utilizados para prever o valor absoluto, *z score* ou classificar a alteração ventilatória presente: regressão linear (LR), regressão logística (LogR), classificador Bayesiano (NB), *k-nearest neighbors* (kNN), *support vector machines* (SVM), redes neurais (NN), florestas de árvores de decisão (RF) e *extreme gradient boosting* (XGboost). Após seleção das variáveis que providenciaram mais informação, os modelos foram submetidos à técnica de reamostragem *20-fold cross validation* e o seu desempenho foi avaliado de acordo com o tipo problema em questão (regressão ou classificação).

Resultados: Na amostra estudada, 66% dos indivíduos apresentaram volumes pulmonares estáticos sem alterações, sendo o *air trapping* a alteração mais prevalente (18,5%).

Para a classificação da presença de *air trapping* e para a classificação da alteração na PCT, o modelo com melhor desempenho foi o XGboost, com uma área abaixo da curva *receiver operating characteristic* (AUC) de 0,881 e 0,874, respetivamente. Em termos de regressão, o algoritmo com melhor desempenho foi a LR, em que na previsão do volume residual (RV) obteve um  $R^2$  de 0,539. Para a capacidade pulmonar total (TLC) o  $R^2$  foi de 0,856 e para o RV/TLC o  $R^2$  foi de 0,752. A previsão do *z score* de RV/TLC apresentou um  $R^2$  de 0,442. O empilhamento de modelos não se mostrou como uma melhor abordagem em comparação com cada modelo utilizado individualmente.

**Conclusão:** O desempenho dos modelos de regressão relativamente ao RV e ao z score de RV/TLC apresentaram resultados insuficientes. No entanto, a regressão para a TLC, para o RV/TLC e a classificação para identificação de *air trapping* e da alteração na PCT tiveram resultados qualitativamente bons, demonstrando a fiabilidade da associação entre a espirometria e a ML na previsão e classificação dos volumes pulmonares não mobilizáveis.

**Palavras-chave:** provas de função respiratória, espirometria forçada, pletismografia corporal total, aprendizagem automática, volumes pulmonares estáticos, hiperinsuflação, restrição.

## Abstract

Background: The most accessible and currently used respiratory function test (PFR) is spirometry, however it is unable to measure static lung volumes, essential to aid in the diagnosis of hyperinflation or lung restriction. Machine learning (ML) techniques have gained notoriety in the medical field largely due to their ability to predict or classify from a large number of already known examples.

Aim: Apply supervised regression models for static lung volumes and automatic classification of their disorders, using biological, anthropometric and spirometric parameters, with evaluation of model's performance.

Methodology: Retrospective study with 8140 anonymized PFRs containing biological, anthropometric, spirometric and total body plethysmography data. Eight models were used to predict the absolute value, z score or classify the ventilatory change: linear regression (LR), logistic regression (LogR), Bayesian classifier (NB), k-nearest neighbors (kNN), support vector machines (SVM), neural networks (NN), decision tree forests (RF) and extreme gradient boosting (XGboost). After selecting the variables that provided more information, the models were submitted to the resampling technique 20-fold cross validation and their performance was evaluated according to the type of problem in question (regression or classification).

Results: In this sample, 66% of the individuals had static lung volumes without alterations, with air trapping being the most prevalent alteration (18.5%).

For the classification of air trapping and for the classification of the plethysmographic disorder, the model with the best performance was the XGboost, with an area under the curve ROC (AUC) of 0.881 and 0.874, respectively. In terms of regression, the algorithm with the best performance was the LR, in which to predict the residual volume (RV) it obtained an  $R^2$  of 0.539. The total lung capacity (TLC) had an  $R^2$  of 0.856 and for the RV/TLC the  $R^2$  was 0.752. The RV/TLC z score prediction had an  $R^2$  of 0.442. Model stacking did not prove to be a better approach compared to each model used individually.

Conclusion The performance of the regression models in relation to the RV and the z score of RV/TLC showed insufficient results. However, the regression for TLC, for RV/TLC and the classification to identify air trapping and the change in PCT had good

results, demonstrating the reliability of the association between spirometry and LM in the prediction and classification of not mobilisable lung volumes.

**Keywords:** respiratory function tests, forced spirometry, whole body plethysmography, machine learning, static lung volumes, hyperinflation, restriction.

# Índice

<b>Resumo</b> .....	iv
<b>Abstract</b> .....	vi
<b>Índice</b> .....	viii
<b>Índice figuras</b> .....	x
<b>Índice tabelas</b> .....	xii
<b>Lista de abreviaturas</b> .....	xiii
<b>1. Introdução</b> .....	1
1.1. Contextualização e problemática.....	1
<b>2. Objetivos</b> .....	2
<b>3. Enquadramento teórico</b> .....	3
3.1. Função Respiratória.....	3
3.1.1. Fisiologia.....	3
3.1.2. Espirometria.....	4
3.1.3. Volumes não mobilizáveis.....	5
3.1.4. Outras técnicas.....	8
3.1.5. Equações de referência.....	9
3.1.6. Interpretação das PFR.....	12
3.2. Aprendizagem automática.....	14
3.2.1. Modelos de Aprendizagem Automática.....	16
3.2.2. Seleção de variáveis e hiperparâmetros.....	21
<b>4. Materiais e Métodos</b> .....	22
4.1. População associada à base de dados.....	22
4.2. Métodos de pré-processamento.....	23
4.3. Estratégias para a análise exploratória dos dados.....	24
4.4. Parâmetros de avaliação de modelos.....	25
<b>5. Resultados</b> .....	30
5.1. Análise exploratória dos dados.....	30
5.2. Modelos.....	39
<b>6. Discussão</b> .....	45

<b>7. Conclusão</b> .....	50
<b>8. Bibliografia</b> .....	52
<b>Anexos</b> .....	59

## Índice figuras

Figura 3.1 – Esquema representativo das curvas volume-tempo e débito-volume obtidas pela espirometria. <sup>7</sup> .....	5
Figura 3.2 - Esquema da PCT e representação dos volumes pulmonares estáticos <sup>2</sup> ...	8
Figura 3.3 – Curva gaussiana da distribuição de uma população normal e correspondentes z-score e probabilidades <sup>18</sup> .....	11
Figura 3.4 - Algoritmo de interpretação para as alterações das vias aéreas, avaliadas por espirometria. <sup>5</sup> .....	13
Figura 3.5 - Algoritmo de interpretação das alterações dos volumes pulmonares estáticos, avaliados por PCT. <sup>5</sup> .....	14
Figura 3.6 - Esquematização de uma <i>5-fold cross validation</i> . <sup>22</sup> .....	16
Figura 3.7 - Esquematização do funcionamento de cada modelo ML utilizado no presente estudo: a) LR <sup>24</sup> ; b) LogR <sup>25</sup> ; c) NB <sup>26</sup> ; d) kNN <sup>27</sup> ; e) SVM <sup>28</sup> ; f) NN <sup>29</sup> ; g) RF <sup>30</sup> ; h) XGboost <sup>31</sup> .....	20
Figura 4.1 - Diagrama de Sankey da depuração da base de dados até à amostra final. ....	23
Figura 4.2 - <i>Boxplot</i> da distribuição da altura, antes e após a remoção de <i>outliers</i> severos. ....	23
Figura 4.3 - Esquema da tabela de contingência, com o significado de cada quadrante e dos cálculos associados para avaliação dos modelos de classificação.....	26
Figura 4.4 - Distribuição de probabilidades e AUC associada <sup>22</sup> .....	27
Figura 5.1 – Histogramas dos z score do FEV <sub>1</sub> , TLC, FEV <sub>1</sub> /FVC e RV com média (z score=0) evidenciada.....	31
Figura 5.2 - Progresso do FEV <sub>1</sub> /FVC, FEV <sub>1</sub> e FVC ao longo da idade. ....	31
Figura 5.3 - Progresso da TLC, VC e RV ao longo da idade. ....	32
Figura 5.4 - Progresso do FEV <sub>1</sub> e do RV ao longo da idade.....	32
Figura 5.5 - Relação entre RV/TLC e FEV <sub>1</sub> (a) e distribuição do FEV <sub>1</sub> pelas categorias da espirometria, divididas pela alteração de RV/TLC (b). ....	33
Figura 5.6 – Distribuição da correlação dos valores absolutos (a) e dos z score (b) por sexo.....	33
Figura 5.7 - <i>Violin plots</i> das categorias da espirometria divididas por sexo, em que se observa uma diminuição das diferenças entre sexos com o agravamento da alteração ventilatória. ....	34
Figura 5.8 - Relação entre o RV/TLC e o IMC (a) e box plots da distribuição de RV/TLC pelas categorias de IMC (b). ....	34

Figura 5.9 - Valores médios de z score do FEV <sub>1</sub> , TLC e RV/TLC de acordo com as categorias de IMC.....	35
Figura 5.10 – Gráficos bidimensionais das variáveis das PFR (a) e heat map com as correlações de Pearson (b).....	36
Figura 5.11 - Tabelas de contingência do XGboost (a), kNN (b) e LogR (c).....	40
Figura 5.12 - Distribuição de probabilidades da LogR (a) e XGboost (b) para RV/TLC normal consoante o z score do FEV <sub>1</sub> com XGboost a apresentar uma maior precisão na identificação de um RV/TLC normal.....	40
Figura 5.13 - Tabela de contingência do modelo de empilhamento. ....	41
Figura 5.14 – <i>Scatter plot</i> entre o valor real e o valor previsto pela LR do z score de RV/TLC, colorida pelas categorias da espirometria, demonstrando um subvalorização das previsões.....	41
Figura 5.15 – <i>Scatter plot</i> entre valor real e o valor previsto pela LR do z score de TLC, colorido de acordo com o FEV <sub>1</sub> pós BD. ....	43
Figura 5.16 - Tabela de contingência do XGboost na classificação da interpretação da PCT. ....	44
Figura 0.1 – <i>Box plots</i> da distribuição dos z score da FVC e FEV <sub>1</sub> /FVC na interpretação da espirometria e da TLC e RV/TLC para a PCT. ....	59
Figura 0.2 - Variação do z score do FEV <sub>1</sub> ao longo da idade dos participantes e de acordo com a gravidade da alteração ventilatória. ....	60
Figura 0.3 - Variação do valor absoluto de RV/LC ao longo da idade dos participantes e de acordo com a gravidade da alteração ventilatória. ....	60
Figura 0.4 - Variação do valor absoluto do FEV <sub>1</sub> ao longo da idade dos participantes e de acordo com a gravidade da alteração ventilatória. ....	60
Figura 0.5 - Esquema utilizado no <i>software Orange</i> para o treino e avaliação do desempenho dos modelos de regressão. ....	61
Figura 0.6 - Esquema utilizado no <i>software Orange</i> para o treino e avaliação do desempenho dos modelos de classificação. ....	61

## Índice tabelas

Tabela 3-1 - Equações de referência mais recentemente publicadas pela ERS/ATS e relevantes para o estudo. <sup>12,17</sup> .....	12
Tabela 3-2 - Hiperparâmetros utilizados em cada um dos modelos. ....	21
Tabela 5-1 - Estatística descritiva univariada das variáveis obtidas das PFR. ....	30
Tabela 5-2 - Distribuição da prevalência das alterações observadas na espirometria. ....	36
Tabela 5-3 - Distribuição da prevalência das alterações observadas na PCT. ....	37
Tabela 5-4 – Tabela cruzada entre as categorias interpretativas da espirometria e da PCT. ....	38
Tabela 5-5 - Tabela cruzada entre as categorias interpretativas da espirometria e da TLCO, com a prevalência da TLCO diminuída por categoria da PCT. ....	39
Tabela 5-6 - Resultados do desempenho dos modelos utilizados na classificação dicotômica de RV/TLC. ....	40
Tabela 5-7 - Resultados do desempenho do empilhamento de modelos na classificação dicotômica de RV/TLC. ....	41
Tabela 5-8 - Resultados do desempenho dos modelos utilizados na regressão do z score de RV/TLC. ....	41
Tabela 5-9 - Resultados do desempenho dos modelos utilizados na regressão do valor absoluto de RV/TLC. ....	42
Tabela 5-10 - Resultados do desempenho dos modelos utilizados na regressão do valor absoluto de RV. ....	41
Tabela 5-11 - Resultados do desempenho dos modelos utilizados na regressão do valor absoluto da TLC. ....	43
Tabela 5-12 - Resultados do desempenho dos modelos utilizados na classificação da interpretação da PCT. ....	43
Tabela 5-13 - Resultados do desempenho do empilhamento de modelos na classificação da interpretação da PCT. ....	43

## Lista de abreviaturas

<b>AUC</b>	<i>Area Under the Curve (ROC)</i>
<b>BD</b>	Broncodilatador
<b>BTPS</b>	<i>Body Temperature, Pressure and water vapor Saturated</i>
<b>DPOC</b>	Doença Pulmonar Obstrutiva Crónica
<b>ERV</b>	<i>Expiratory Reserve Volume</i>
<b>ERS/ATS</b>	<i>European Respiratory Society/American Thoracic Society</i>
<b>FEV<sub>1</sub></b>	<i>Forced Expiratory Volume in the first second</i>
<b>FRC</b>	<i>Functional Residual Capacity</i>
<b>FVC</b>	<i>Forced Vital Capacity</i>
<b>HLL</b>	Hospital da Luz de Lisboa
<b>IC</b>	<i>Inspiratory Capacity</i>
<b>IRV</b>	<i>Inspiratory Reserve Volume</i>
<b>IMC</b>	Índice de Massa Corporal
<b>kNN</b>	<i>k Nearest Neighbors</i>
<b>LLN</b>	<i>Lower Limit of Normal</i>
<b>LogR</b>	<i>Logistic Regression</i>
<b>LR</b>	<i>Linear Regression</i>
<b>MAE</b>	<i>Mean Absolute Error</i>
<b>ML</b>	<i>Machine Learning</i>
<b>MSE</b>	<i>Mean Square Error</i>
<b>NB</b>	<i>Naive Bayes</i>
<b>NN</b>	<i>Neural Networks</i>
<b>PCT</b>	Pletismografia Corporal Total
<b>RFT</b>	Provas de Função Respiratória
<b>RF</b>	<i>Random Forests</i>
<b>RMSE</b>	<i>Root of Mean Square Error</i>
<b>ROC</b>	<i>Receiver Operating Curve</i>
<b>RV</b>	<i>Residual Volume</i>
<b>SI</b>	Sistema Internacional

<b>SVM</b>	<i>Support Vector Machines</i>
<b>TLC</b>	<i>Total Lung Capacity</i>
<b>TLCO</b>	<i>Transfer Factor of the Lung for Carbon Monoxide</i>
<b>TV</b>	<i>Tidal Volume</i>
<b>ULN</b>	<i>Upper Limit of Normal</i>
<b>XGboost</b>	<i>Extreme Gradient boosting</i>

# 1. Introdução

## 1.1. Contextualização e problemática

Nos últimos 30 anos verificou-se um aumento de 40% da incidência de doenças respiratórias a nível mundial. Estima-se que atualmente existam 545 milhões de doentes respiratórios.<sup>1</sup> O seguimento das doenças respiratórias depende, não exclusivamente, das provas de função respiratória (PFR), das quais a espirometria forçada (comumente denominada por espirometria) é a mais utilizada e acessível, dado que é um procedimento independente de laboratórios especializados.

A espirometria fornece informação sobre os volumes pulmonares dinâmicos (volumes mobilizáveis que necessitam de manobras forçadas para serem gerados) e débitos expiratórios e inspiratórios, no entanto é insuficiente para medir a quantidade de ar que permanece no pulmão após uma expiração completa: o volume residual (RV). O RV é um volume estático fundamental pois é imprescindível para se conhecer a quantidade máxima de ar que o pulmão é capaz de reter numa manobra inspiratória máxima, denominada por capacidade pulmonar total (TLC). Estes dois volumes pulmonares não mobilizáveis são essenciais para a identificação da hiperinsuflação pulmonar ou para a confirmação da presença de alterações ventilatórias restritivas.<sup>2</sup>

Contudo, a obtenção destes dados só é possível através de outros exames e com recurso a outros equipamentos. Assim, a medição dos volumes não mobilizáveis restringe-se a laboratórios especializados e pouco acessíveis à população geral.

As técnicas de aprendizagem automática (ML) utilizam sistemas computacionais de forma a aprenderem sem a necessidade de instruções específicas, possibilitando inferir informação a partir de um elevado número de dados de qualidade.

A simbiose entre ML e PFR aparenta ter crescente interesse na comunidade científica, com vários estudos realizados que se focaram na melhoria do diagnóstico das patologias respiratórias e na previsão da sua progressão ao longo do tempo.<sup>3,4</sup>

Face ao exposto, considerou-se a ML como uma ferramenta com potencial para acrescentar informação obtida pela espirometria, através do desenvolvimento de metodologias de previsão do RV e da TLC com base em parâmetros espirométricos.

## 2. Objetivos

Os objetivos principais do estudo foram:

- Aplicar modelos de ML supervisionados de classificação para a deteção automática do aumento categórico do RV/TLC (*air trapping*) e avaliar o seu desempenho.
- Aplicar modelos de ML supervisionados preditivos de regressão para o RV, a TLC e a sua relação (RV/TLC) e avaliar o seu desempenho;

No sentido de verificar de que forma é que as características dos participantes influenciaram os modelos preditivos de ML aplicados, definiram-se os seguintes objetivos específicos:

- a) Correlacionar as características biológicas, antropométricas e os volumes pulmonares;
- b) Identificar o contributo da resposta ventilatória ao broncodilatador para a previsão dos volumes pulmonares não mobilizáveis.

## 3. Enquadramento teórico

Este capítulo é composto por um aprofundamento teórico relacionado com as duas principais áreas que sustentam esta dissertação: o estudo da função respiratória e os modelos de aprendizagem automática.

### 3.1. Função Respiratória

#### 3.1.1. Fisiologia

As PFR refletem as propriedades fisiológicas do aparelho respiratório e têm sido utilizadas há décadas no apoio ao diagnóstico da patologia pulmonar, no esclarecimento da dispneia e na monitorização da progressão e resposta ao tratamento deste tipo de doenças.<sup>5</sup>

Para compreender e interpretar convenientemente as PFR é necessário um conhecimento das interações estruturais e funcionais do sistema respiratório. A respiração é modulada por centros respiratórios que, mediante a informação recolhida pelos quimiorreceptores centrais e periféricos, fornecem aos músculos respiratórios informação para que estes variem a frequência e o volume da ventilação. A contração destes músculos gera o gradiente de pressão necessário para a inspiração, que conduz o ar das zonas de condução da árvore brônquica para os alvéolos, onde as trocas gasosas ocorrem. O seu relaxamento, associado à pressão de retração elástica do pulmão, permite o gradiente necessário à expiração. Através da circulação sanguínea pulmonar, o sangue desoxigenado e rico em metabolitos (entre os quais o dióxido de carbono – CO<sub>2</sub>) entra em contacto com os alvéolos. É nesta fase que ocorrem as trocas gasosas, especificamente a difusão de gases através da membrana alvéolo-capilar. O oxigénio (O<sub>2</sub>) alveolar difunde-se para o sangue e liga-se à hemoglobina presente nos eritrócitos e o CO<sub>2</sub> passa para o ar alveolar e é por sua vez expirado para o ar atmosférico.<sup>6</sup>

A doença pulmonar pode ocorrer em qualquer fase deste complexo processo fisiológico e as PFR, com a sua capacidade de evidenciar a fisiologia do sistema respiratório, têm um papel essencial para a gestão destas patologias, nomeadamente no diagnóstico, na avaliação da progressão e na resposta à terapêutica.<sup>6</sup>

### 3.1.2. Espirometria

A espirometria é um exame que permite medir volumes pulmonares mobilizáveis e débitos das vias aéreas. Pode ser lenta ou forçada consoante interesse (ou não) a velocidade com que o ar é mobilizado. Na espirometria forçada os parâmetros são recolhidos no seguimento de uma expiração forçada, rápida e completa, após uma inspiração máxima. O volume total de ar medido nesta manobra é denominado de capacidade vital forçada ou *Forced Vital Capacity* (FVC), que se inicia, em termos de volumes não mobilizáveis, na capacidade pulmonar total ou *Total Lung Capacity* (TLC) e termina no volume residual ou *Residual Volume* (RV).<sup>2,7-9</sup>

Vários parâmetros volumétricos determinantes da saúde respiratória de um indivíduo podem ser recolhidos através desta manobra, no entanto, pode considerar-se como principal o volume expiratório máximo no primeiro segundo ou *Forced Expiratory Volume in the first second* (FEV<sub>1</sub>). Todos os volumes recolhidos são corrigidos para BTPS (*body temperature, pressure and water vapor saturated*) devido às diferenças de temperatura e saturação de vapor de água entre o ar inspirado e expirado.<sup>2,7-9</sup>

A relação entre o FEV<sub>1</sub> e FVC (FEV<sub>1</sub>/FVC) é um índice que revela a existência da diminuição do calibre das vias aéreas e conseqüentemente uma obstrução da passagem do ar através das mesmas. Esta relação é essencial para o diagnóstico de algumas patologias pulmonares, entre as quais a Doença Pulmonar Obstrutiva Crónica (DPOC). Por exemplo, um FEV<sub>1</sub>/FVC de 0,6 indica que, na manobra expiratória máxima, 60% do volume de ar foi expirado no primeiro segundo (FEV<sub>1</sub>) e os restantes 40% no tempo remanescente.<sup>2,7-9</sup>

Através da espirometria é ainda possível medir débitos expiratórios com importância fisiopatológica, mas que não serão abordados por não serem relevantes no contexto desta dissertação.

As alterações ventilatórias identificadas pela espirometria designam-se por obstrutivas, restritivas ou mistas. No entanto, não é possível aferir com certeza se uma diminuição da FVC reflete uma diminuição da TLC ou um aumento do RV.<sup>2,7-9</sup> Desta limitação, surgem os objetivos principais da atual dissertação.

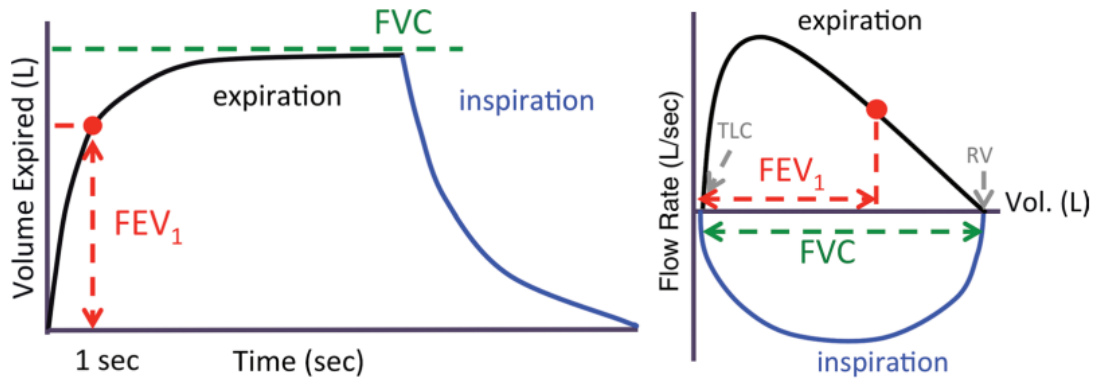


Figura 3.1 – Esquema representativo das curvas volume-tempo e débito-volume obtidas pela espirometria.<sup>7</sup>

### 3.1.3. Volumes não mobilizáveis

Os volumes pulmonares não mobilizáveis são considerados aqueles que não podem ser medidos diretamente pelos sensores de débito através de manobras ventilatórias.<sup>2,9</sup>

Por convenção, os volumes pulmonares estáticos estão divididos em volumes e capacidades (conjuntos de dois ou mais volumes). Assim, quatro volumes pulmonares dão origem a quatro capacidades, como no esquema da figura 3.2.<sup>2,10-12</sup>

$$\rightarrow FRC = RV + ERV$$

$$\rightarrow IC = VT + IRV$$

$$\rightarrow TLC = RV + ERV + VT + IRV$$

$$\rightarrow VC = ERV + VT + IRV$$

Nestas equações, *IC* é a capacidade inspiratória ou *Inspiratory Capacity*, *ERV* é o volume de reserva expiratória ou *Expiratory Reserve Volume* e *IRV* é o volume de reserva inspiratória ou *Inspiratory Reserve Volume*.

A capacidade residual funcional ou *Functional Reserve Capacity (FRC)* é a quantidade de ar que permanece no pulmão no final da expiração a volume corrente e representa o equilíbrio entre as forças de expansão da caixa torácica e de retração elástica do pulmão.<sup>9</sup>

Na espirometria forçada, o volume total mobilizado é denominado por FVC, no caso da espirometria lenta, esse volume denomina-se capacidade vital ou *Vital Capacity (VC)*. A FVC e a VC são idênticas em indivíduos saudáveis, no entanto, em patologias em que

se verifica o colapso das vias aéreas (nomeadamente doenças pulmonares obstrutivas graves), a VC é tendencialmente mais elevada do que a FVC.<sup>10</sup>

Como já referido, a medição do RV e da TLC é relevante para o diagnóstico, a caracterização e a monitorização de patologias associadas a restrição (diminuição da TLC) ou hiperinsuflação (aumento de TLC e/ou RV/TLC) pulmonar.<sup>2,10,13</sup>

Para a medição destes volumes existem três métodos principais: a diluição de hélio (ou outro gás insolúvel), a depuração (*wash-out*) de azoto e a pletismografia corporal total (PCT). Os dois primeiros métodos baseiam-se na análise de gases em concentrações que não estão presentes habitualmente no pulmão e não serão aprofundados.<sup>2,12</sup>

A PCT é o *gold standard* para a medição dos volumes estáticos e mede diretamente a capacidade residual funcional ( $FRC_{pleth}$ ) baseando-se na lei de Boyle-Mariotte, segundo a qual o volume e a pressão variam de forma inversamente proporcional em condições isotérmicas. Assim, o produto entre volume e pressão é uma constante.<sup>11,13</sup>

$$P \times V = k$$

Para a medição da  $FRC_{pleth}$ , a PCT é uma técnica que utiliza uma cabine hermeticamente fechada (volume constante) com transdutores que medem a pressão na cabine ( $P_{cabine}$ ) e na boca do indivíduo ( $P_{boca}$ ).<sup>11</sup> A  $FRC_{pleth}$  obtém-se através da fórmula:

$$P_1 \times V_1 = (P_1 - \Delta P) \times (V_1 + \Delta V)$$

Nesta equação,  $P_1$  é a pressão barométrica medida no final da expiração ( $P_{atmosférica} - P_{Vapor H_2O}$ ), quando o seu valor é idêntico no pulmão, na boca e na cabine,  $V_1$  é a incógnita ( $FRC_{pleth}$ ),  $\Delta P$  é a variação da pressão medida no sensor da boca e  $\Delta V$  é a variação do gás intratorácico.<sup>11</sup>

Assim, quando o indivíduo está a respirar a volume corrente, uma válvula encerra no final da expiração. Este é instruído a continuar a realizar o esforço ventilatório a uma dada frequência, que leva a que o ar que está nos pulmões seja comprimido e descomprimido. Durante esse período em que não há débito aéreo, as variações de pressão que ocorrem na boca consideram-se idênticas às variações de pressão que ocorrem no pulmão ( $\Delta P_{alveolar} = \Delta P_{boca}$ ). Simultaneamente está a ser lida a variação

de pressão na cabine que resulta dos movimentos torácicos e que reflete a variação do volume intrapulmonar. <sup>11</sup>

Assim, conhecendo o volume da cabine, a variação de pressão da cabine e da boca e ignorando a  $P_{atmosférica}$ , obtemos a seguinte fórmula para o cálculo da  $FRC_{pleth}$ .<sup>11</sup>

$$FRC_{pleth} = -V_{cabine} \times \frac{\Delta P_{cabine}}{\Delta P_{boca}}$$

O sinal negativo não irá gerar resultados paradoxais (volume negativo) pois a variação da pressão na cabine e da variação da pressão na boca terão sinais opostos, pelo que o resultado será sempre positivo. <sup>11</sup>

Quando a válvula abre, é realizada uma VC de modo a obter-se os volumes mobilizáveis que permitirão determinar a TLC e o RV. <sup>11</sup>

$$\rightarrow TLC = FRC + IC$$

$$\rightarrow RV = TLC - VC$$

Ou

$$\rightarrow RV = FRC - ERV$$

$$\rightarrow TLC = RV + VC$$

A TLC e o RV são determinados por diferentes conjuntos musculares, que atuam em diferentes estruturas. A TLC é sensível a variações das propriedades da caixa torácica, dos músculos inspiratórios e do próprio pulmão enquanto o RV é sensível a variações das propriedades dos músculos expiratórios e a mecanismos de limitação dos débitos expiratórios. <sup>10</sup>

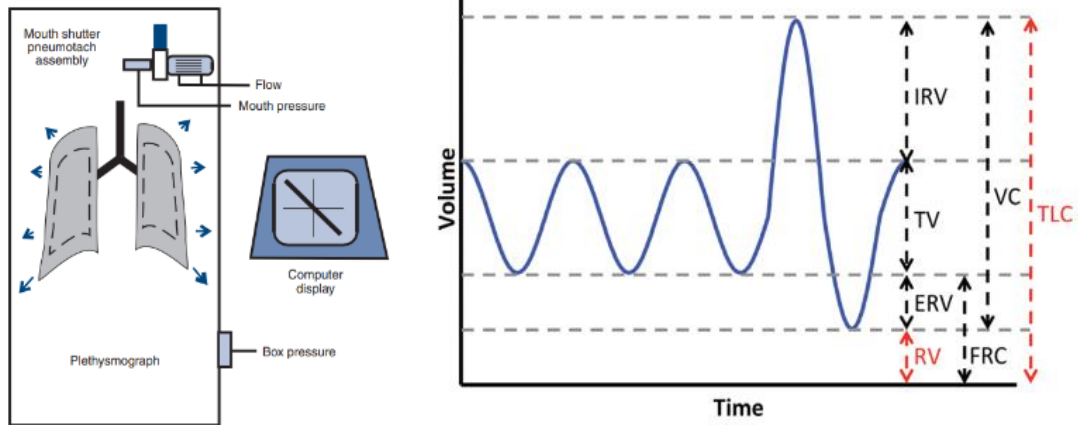


Figura 3.2 - Esquema da PCT e representação dos volumes pulmonares estáticos<sup>2</sup>

Frequentemente, patologias extrapulmonares refletem-se na mecânica ventilatória, como é o caso das alterações da caixa torácica ou doenças neuromusculares.<sup>5</sup>

A obesidade também é uma patologia que se reflete na mecânica ventilatória, justificando limitações do fluxo expiratório ou hipoventilação alveolar. Em termos de volumes pulmonares, é possível observar uma diminuição marcada do ERV, por diminuição de FRC e aumento de RV. Em casos mais graves, com Índice de Massa Corporal (IMC) superior a 40 kg.m<sup>-2</sup>, a limitação à expansão gerada na caixa torácica é de tal ordem que pode causar alterações restritivas com diminuição da TLC.<sup>5,14</sup>

### 3.1.4. Outras técnicas

Existem várias técnicas e provas que podem ser realizadas de forma a avaliar características específicas da função respiratória. Duas técnicas serão abordadas de forma superficial nesta dissertação: a prova de resposta ao broncodilatador (BD) e o estudo da transferência alvéolo-capilar do monóxido de carbono (TLCO).<sup>15,16</sup>

O intuito da prova de resposta ao BD é avaliar, de forma mensurável e objetiva, a variabilidade da permeabilidade das vias aéreas na presença de um fármaco com características broncodilatadoras, ou seja, capaz de relaxar o músculo liso presente maioritariamente nas porções médio-terminais da árvore brônquica, permitindo assim o aumento do diâmetro do lúmen das vias aéreas, que por sua vez se traduzirá no aumento dos débitos expiratórios e dos volumes mobilizáveis. Uma resposta positiva ao BD traduz-se numa variação específica do FEV<sub>1</sub> e/ou da FVC.

$$BDR = \frac{(PostBD - PreBD)}{Valor Previsto} \times 100$$

Em que *BDR* é o valor da resposta ao broncodilatador, *PostBD* é valor medido ( $FEV_1$  ou FVC) após a administração do BD, *PreBD* valor medido ( $FEV_1$  ou FVC) antes da administração do BD e o *Valor Previsto* é a previsão de acordo com o sexo, idade, altura e raça de uma população considerada saudável. Se *BDR* for superior a 10%, a resposta é considerada como positiva.<sup>5</sup>

Contudo valorizam-se também outras alterações em diversos parâmetros, tais como a diminuição da resistência das vias aéreas, o aumento dos débitos expiratórios ou a diminuição da hiperinsuflação pulmonar por redistribuição de volumes não mobilizáveis (aumento da VC/FVC com conseqüente diminuição do RV). É uma prova bastante acessível e que se associa frequentemente à espirometria, pelo que pode ter potencial para gerar informação útil para os modelos de ML.<sup>2,8</sup>

A TLCO por respiração única revela a capacidade de difusão gasosa da barreira alveo-capilar de um dado sistema respiratório. A técnica é realizada através da inalação de dois gases com concentrações específicas: um gás insolúvel (ou *tracer gas*) e o monóxido de carbono (CO), que apresenta uma grande afinidade para a hemoglobina e um comportamento semelhante ao  $O_2$  no que diz respeito à difusão alvéolo-capilar.<sup>15</sup> É uma técnica utilizada maioritariamente em laboratórios especializados, pelo que o seu propósito neste estudo será apenas de caracterização da população, não sendo incluído no treino de nenhum modelo de ML.

### 3.1.5. Equações de referência

Ao contrário da maioria dos índices biológicos usados em Medicina, como as concentrações plasmáticas de eletrólitos ou de algumas hormonas, as PFR variam significativamente com a idade, a altura, a raça e o sexo.<sup>5,12,17</sup>

A idade tem dois contributos importantes para a função respiratória: o crescimento somático está fortemente ligado ao volume da caixa torácica e ao calibre das vias aéreas (com exceção os períodos de crescimento rápido, em que se observa uma assincronia nesta relação) e o processo de envelhecimento natural nos adultos leva a um aumento da rigidez da caixa torácica, atrofia dos músculos respiratórios e perda de elasticidade do pulmão. A altura por si só não é um determinante direto dos volumes pulmonares, no entanto é um indicador das dimensões do tórax (assim como a raça). O sexo é um indicador importante do tamanho do pulmão, pois mesmo com

alturas semelhantes, indivíduos do sexo masculino têm, em média, valores mais elevados do que os do sexo feminino. <sup>5</sup>

Um parâmetro com o mesmo valor numérico, poderá ser normal ou não normal, dependendo das características identificadas anteriormente. De forma a facilitar a interpretação das PFR, foram introduzidos limites inferiores (LLN) e superiores da normalidade (ULN), de acordo com um padrão estatístico denominado por *z score*. <sup>5</sup>

Para as PFR, definiu-se como limites da normalidade o 5<sup>o</sup> e o 95<sup>o</sup> percentil (correspondendo aos *z score* -1,645 e +1,645, respetivamente) de uma população saudável. Assim, um valor que esteja contido neste intervalo apresenta 90% de probabilidade de fazer parte da população considerada saudável (figura 3.3). O *z score* igual a zero é a média da população normal e cada unidade representa um desvio-padrão (que tanto pode ter sinal negativo como positivo, dependendo da sua posição relativa à média). <sup>5</sup>

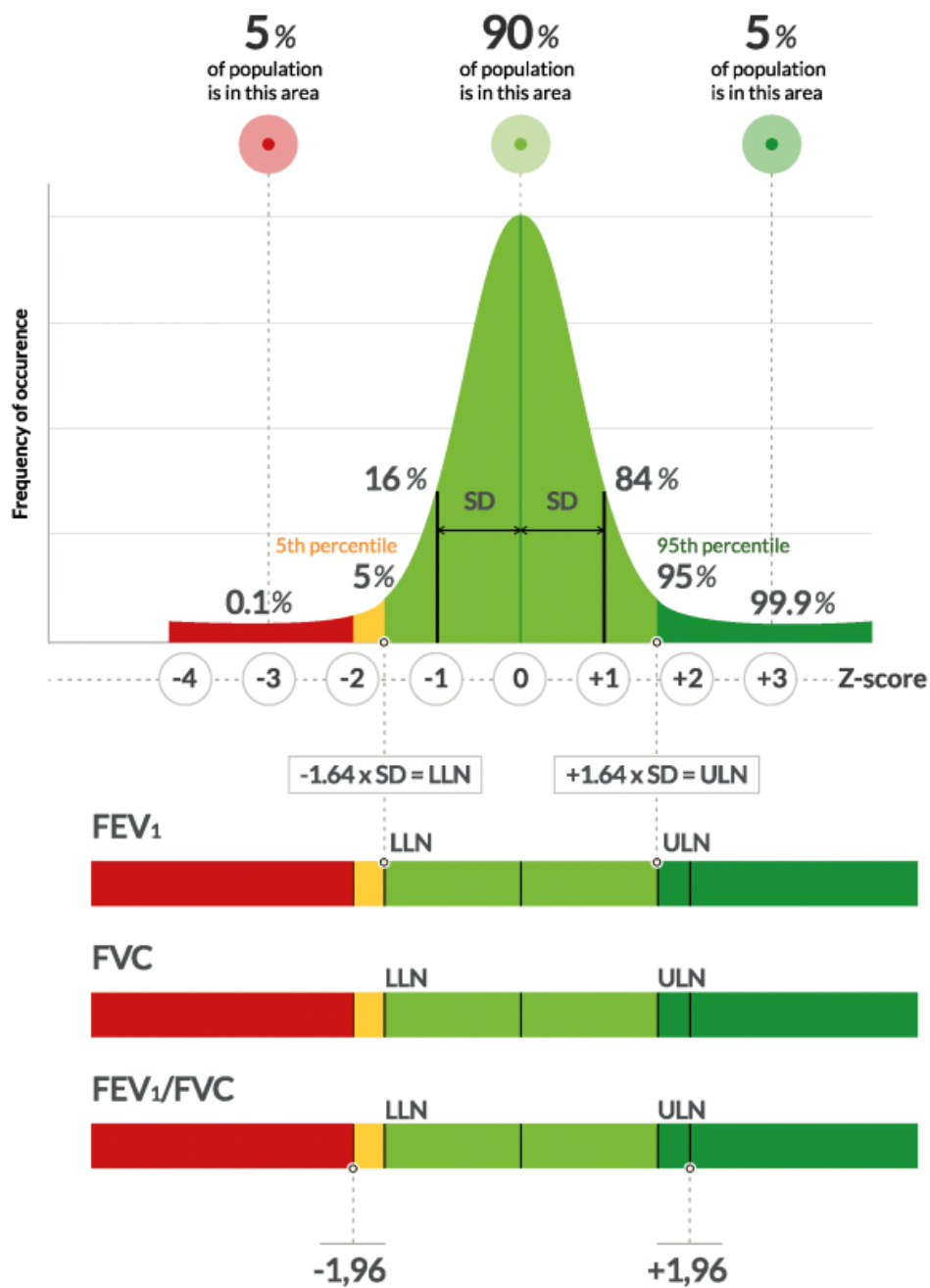


Figura 3.3 – Curva gaussiana da distribuição de uma população normal e correspondentes z-scores e probabilidades.<sup>18</sup>

Este parâmetro estatístico é uma abordagem diferente da percentagem do valor previsto, calculada segundo a seguinte equação: <sup>5</sup>

$$\% \text{ Previsto} = \frac{\text{Valor medido}}{\text{Valor previsto}} \times 100$$

Devido às características fisiológicas associadas à mecânica ventilatória e à transferência alvéolo-capilar, alguns valores fora dos limites da normalidade podem não

estar necessariamente alterados. Por exemplo, no caso dos volumes dinâmicos medidos por espirometria forçada, um valor anormalmente alto apenas representa uma excelente capacidade de mobilizar o ar por parte desse indivíduo. <sup>5</sup>

De forma a avaliar e a comparar a função respiratória de um dado indivíduo em relação à normalidade, ao longo dos anos foram criadas várias equações de referência baseadas em exames realizados numa população considerada saudável, estando as equações mais recentes para a população europeia expostas na tabela 3.1.

Tabela 3-1 - Equações de referência mais recentemente publicadas pela ERS/ATS e relevantes para o estudo. <sup>12,17</sup>

Ano	PFR	Parâmetros (adultos)	Intervalo etário (anos)	LLN	ULN
2012	<b>Espirometria(18)</b>	FEV <sub>1</sub> , FVC, FEV <sub>1</sub> /FVC, FEF <sub>25-75%</sub>	3 a 95	✓	✗
2021	<b>Volumes estáticos(13)</b>	FRC, TLC, RV, ERV, IC, VC, RV/TLC	5 a 80	✓	✓

### 3.1.6. Interpretação das PFR

Em 2022 foram publicadas pela *European Respiratory Society* e pela *American Thoracic Society* (ERS/ATS), novas estratégias interpretativas para as PFR de rotina: a espirometria, os volumes pulmonares estáticos (PCT) e a TLCO. <sup>2,7,8</sup>

As alterações das PFR estão subdivididas em três grupos: alterações das vias aéreas, dos volumes pulmonares estáticos e/ou da transferência alvéolo-capilar. <sup>5</sup>

As alterações presentes nas vias aéreas são avaliadas de acordo com parâmetros medidos pela espirometria. <sup>5</sup>

- Normal: FEV<sub>1</sub>/FVC superior ao LLN e FVC superior ao LLN.
- Obstrução: FEV<sub>1</sub>/FVC inferior ao LLN e FVC superior ao LLN;
- Restrição: FEV<sub>1</sub>/FVC superior ao LLN e FVC inferior ao LLN;
- Misto: FEV<sub>1</sub>/FVC inferior ao LLN e FVC inferior ao LLN;

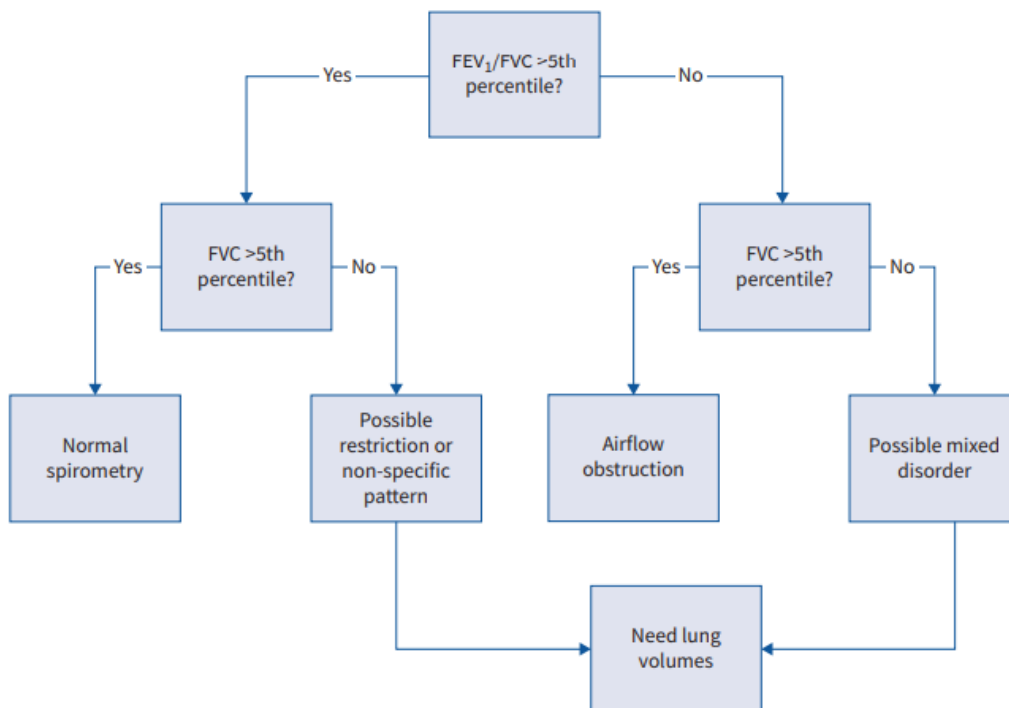


Figura 3.4 - Algoritmo de interpretação para as alterações das vias aéreas, avaliadas por espirometria. <sup>5</sup>

As alterações presentes nos volumes pulmonares estáticos são avaliadas de acordo com parâmetros medidos através de PCT e espirometria lenta. <sup>5</sup>

- Hiperinsuflação: TLC superior ao ULN e RV/TLC superior ao ULN;
- *Air trapping* (subcategoria da hiperinsuflação): TLC entre LLN e ULN e RV/TLC superior ao ULN;
- Volumes proporcionalmente aumentados: TLC superior ao ULN e RV/TLC inferior ao ULN;
- Volumes pulmonares normais: TLC entre LLN e ULN e RV/TLC inferior ao ULN;
- Restrição simples: TLC inferior ao LLN, RV/TLC inferior ao ULN e FEV<sub>1</sub>/FVC superior ao LLN;
- Restrição complexa: TLC inferior ao LLN, RV/TLC superior ao ULN e FEV<sub>1</sub>/FVC superior ao LLN;
- Mista: TLC inferior ao LLN e FEV<sub>1</sub>/FVC inferior ao LLN.

A diferença entre hiperinsuflação e *air trapping* prende-se com o facto de que, na primeira situação, o aumento da TLC representa uma alteração estrutural do pulmão, com perda de retração elástica e é um fenómeno associado ao enfisema. No caso do *air trapping*, fenómeno observado principalmente na asma brônquica ou na bronquite crónica, a limitação ao fluxo expiratório é responsável pelo encerramento prematuro das

vias aéreas, provocando um aumento do RV, mantendo a TLC inalterada. Desta forma, na interpretação das PFR o aumento do RV/TLC é mais valorizado do que o aumento isolado do RV ou da TLC.<sup>5</sup>

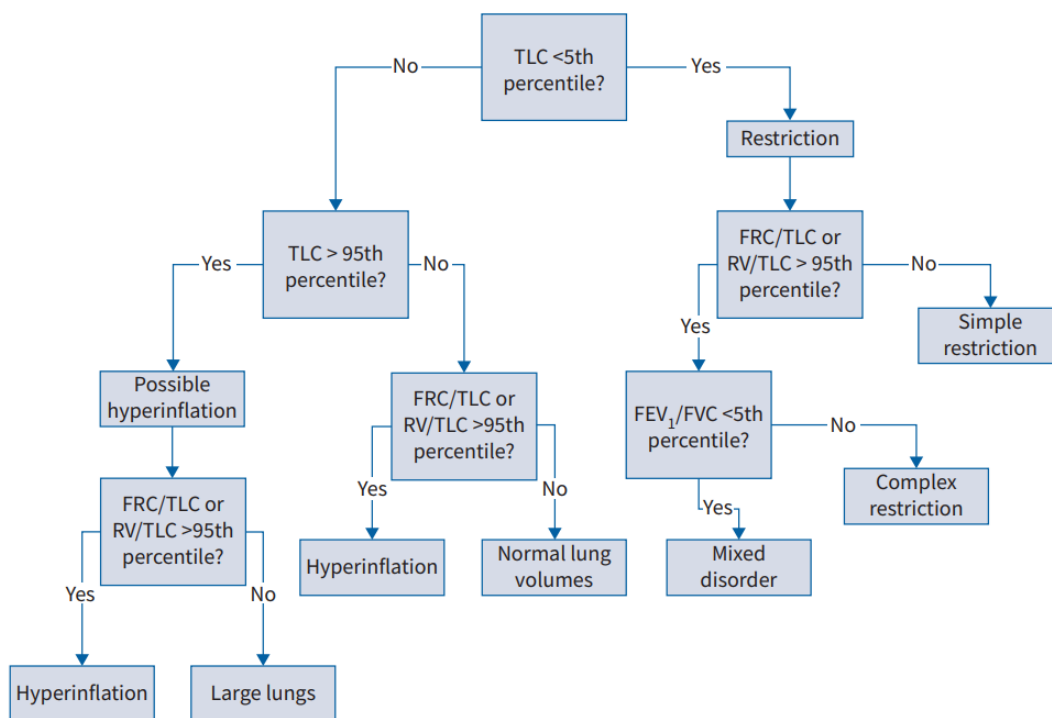


Figura 3.5 - Algoritmo de interpretação das alterações dos volumes pulmonares estáticos, avaliados por PCT.<sup>5</sup>

A TLCO é considerada como alterada se o z score for inferior a -1,645 e normal se o z score for superior a -1,645.<sup>5</sup>

### 3.2. Aprendizagem automática

A aprendizagem automática ou *machine learning* (ML), é um subconjunto de Inteligência Artificial que, apesar de não ser recente, adquiriu relevância na comunidade científica, devido aos avanços computacionais e ao crescimento exponencial de dados disponíveis.<sup>3,4</sup>

Entre outras áreas, a Medicina pode beneficiar deste tipo de abordagem, pois a ML permite a análise de dados multidimensionais e multiparamétricos revelando complexos padrões nos dados clínicos, permitindo um eficiente apoio à decisão clínica e uma melhoria na prestação de cuidados de saúde.<sup>3,4</sup>

Existem duas grandes classes de modelos (ou algoritmos) de ML: os modelos supervisionados e os não supervisionados. <sup>19-22</sup>

Os modelos não supervisionados são utilizados para identificar padrões nos dados previamente desconhecidos, inferindo resultados de informação não catalogada. Tipicamente adotam uma de duas abordagens: o estudo de características de similaridades entre os dados ou o estudo das diferenças (*outliers*) sobre a maioria. <sup>19-22</sup>

Por sua vez, os modelos supervisionados aprendem com dados previamente catalogados e, dessa aprendizagem, ganham a capacidade de prever ou classificar autonomamente amostras de dados teste (nunca utilizadas). Esta dissertação recairá neste tipo de modelos, pelo que se irá de seguida detalhar os diferentes modelos e estratégias de aplicação. <sup>19-22</sup>

Os modelos supervisionados são divididos em modelos de classificação ou de regressão dependendo do tipo de alvo (*target*) que se pretende prever ou classificar. <sup>19-</sup>

22

Neste sentido, os modelos de classificação permitem categorizar de forma autónoma alvos de classes (por exemplo, classificar presença ou ausência de uma dada patologia) e os modelos de regressão permitem prever um valor numérico a partir de outras variáveis (por exemplo, prever a concentração de uma dada hormona no sangue).

Todos os modelos supervisionados de ML necessitam de treinar num conjunto de dados (*train set*) e depois são avaliados num conjunto de dados independente da amostra de treino (*test set*). Existem várias metodologias para esta divisão, sendo uma das mais comuns a divisão fixa da amostra (por exemplo relação *train set/test set* de 80/20). Habitualmente existe uma terceira entidade na qual os hiperparâmetros dos modelos são otimizados, chamada de amostra de validação (*validation set*). Esta divisão do conjunto de dados em estudo é recomendada, exceto se a quantidade de dados for escassa. Deve utilizar-se *k-fold cross validation* se a amostra for inferior a 1000 indivíduos. Entre 1000 e 10 000, a decisão de qual a metodologia dependerá do investigador e dos dados em questão. <sup>21</sup>

O *Cross validation* (CV) é um procedimento de reamostragem que permite estimar qual o erro de um modelo. Nesta metodologia, a amostra inicial é dividida em várias mais pequenas dependendo do valor *k*, por exemplo, com um *k=20*, cada amostra ou *fold* será 1/20 da população inicial. De seguida, uma *fold* é separada das restantes e

o modelo treina nas restantes 19  *folds*. A  *fold* deixada de parte é depois utilizada como amostra de teste. Este procedimento é repetido várias vezes e o resultado a avaliar é média de todos os valores computados no procedimento, como é possível observar na figura 3.6. <sup>21</sup>

Este procedimento pode ser mais dispendioso em termos computacionais, no entanto, com a evolução tecnológica esta é cada vez menos uma desvantagem. <sup>21</sup>

A CV deve ser estratificada, ou seja, cada  *fold* terá características idênticas entre si com as mesmas proporções de classes que a amostra completa. <sup>21</sup>

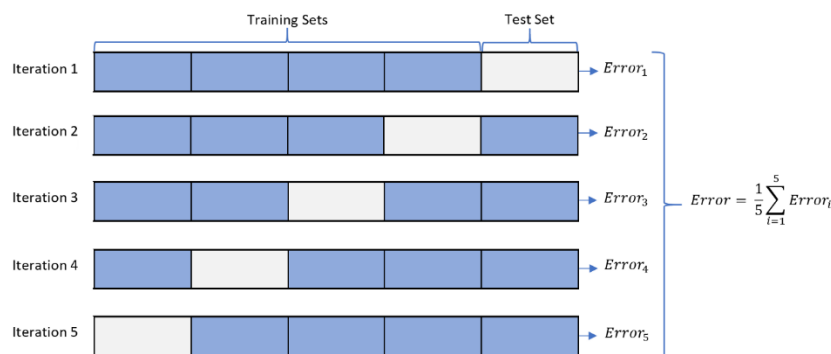


Figura 3.6 - Esquematização de uma 5-fold cross validation.<sup>23</sup>

### 3.2.1. Modelos de Aprendizagem Automática

Cada modelo de ML apresenta características próprias, que poderão ditar o sucesso ou insucesso da predição dos seus resultados, de acordo com o tipo de dados que lhe são fornecidos. Nesta seção da dissertação, será feita uma breve introdução ao funcionamento inerente de cada tipo de algoritmo, com esquematização na figura 3.7.

19–22,24

→ Linear Regression (LR)

A regressão linear (*linear regression*) é um método simples e poderoso para resolver problemas de regressão. A sua função é definida por:

$$y = \vec{\beta}_0 + \vec{\beta}_1x_1 + \vec{\beta}_2x_2 + \dots + \vec{\beta}_nx_n$$

Na função referida,  $y$  é a variável alvo,  $x$  são as variáveis utilizadas como *input* e  $\vec{\beta}$  são os coeficientes de cada variável. Os coeficientes são importantes pois irão revelar a importância associada a cada variável.

Apesar das suas vantagens de simplicidade e interpretabilidade, as LR não podem ser utilizadas em problemas de classificação ou quando as relações entre os dados são complexas e não lineares.

→ Logistic Regression (LogR)

Outro algoritmo bastante popular devido à sua simplicidade é a regressão logística (*logistic regression*), cuja probabilidade de uma dada classe alvo é calculada pela função:

$$P(y|x_1, x_2, \dots, x_n) = \frac{1}{1 + e^{-(\vec{\beta}_0 + \vec{\beta}_1 x_1 + \vec{\beta}_2 x_2 + \dots + \vec{\beta}_n x_n)}}$$

Nesta função, tal como na LR,  $y$  é a variável alvo,  $x$  são as variáveis utilizadas como *input* e  $\vec{\beta}$  são os coeficientes de cada variável.

Esta função tem como resultado um valor entre 0 e 1, que representa a probabilidade de o *output* gerado pertencer à classe positiva. Este algoritmo é puramente classificativo e não pode ser utilizado em problemas de regressão.

→ Naïve Bayes (NB)

Este algoritmo classificativo, também conhecido por classificador Bayesiano, baseia-se no teorema de Bayes, que descreve a probabilidade de um evento ocorrer de acordo com o conhecimento prévio das condições associadas a esse evento:

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)}$$

Da equação temos  $P(A|B)$  que é a probabilidade de a hipótese estar correta dado que a evidência é verdadeira, ou seja, a crença de que modelo acerta na classificação real de um indivíduo aleatório.  $P(B|A)$  é a probabilidade de a evidência ser verdadeira se a hipótese for verdadeira.  $P(A)$  e  $P(B)$  são as probabilidades *a priori* da evidência e da hipótese, respetivamente, ou seja, as suas prevalências.

Este classificador assume a independência de cada variável em relação às restantes, considerando que cada variável contribui de forma igual e independente para o resultado.

Apesar de ser de simples utilização e necessitar de quantidades relativamente baixas de dados para gerar probabilidades fidedignas, as variáveis utilizadas como

*input* com correlação entre si irão contribuir negativamente para o desempenho do modelo e este assume que as variáveis numéricas apresentam distribuição normal.

→ K-nearest Neighbours (kNN)

Quando este algoritmo é utilizado, assume que os dados aos quais nunca esteve exposto são designados de acordo com a média/mediana (em problemas de regressão) ou com a moda (problemas de classificação) das  $k$  amostras mais semelhantes encontradas na amostra de treino, em que  $k$  é um número inteiro, não nulo.

Apesar de apresentar resultados rapidamente e poder ser utilizado para problemas de regressão ou de classificação, necessita de mais memória e tempo de processamento computacional quanto maior for a dimensão da amostra, sendo que, o desempenho do modelo é exponencialmente melhor quantas mais variáveis são utilizadas.

→ Support Vector Machines (SVM)

Este algoritmo de classificação binário identifica as classes de *output* definindo um limite ótimo de decisão, chamado de hiperplano, no espaço que melhor separa as classes. Em teoria, qualquer número de hiperplanos pode ser usado.

Desta forma, o algoritmo apenas utiliza dados à volta do hiperplano, os chamados *support vectors*, maximizando as suas margens de forma que cada novo dado seja classificado de acordo com a sua distância relativa aos limites de decisão.

Assim, este modelo apresenta pouco risco de *overfitting* (classificação demasiado sensível e que responde mal quando encontra novos dados dos quais não tinha conhecimento prévio) e necessita de relativamente poucos dados para obter bons desempenhos. Por outro lado, não obtém bons desempenhos em grandes amostras de dados complexos e/ou com demasiado ruído (*outliers*).

→ Neural Networks (NN)

As redes neuronais ou *neural networks* são conjuntos de algoritmos cujo funcionamento é inspirado nas redes biológicas de neurónios do cérebro humano. No cérebro humano, os neurónios estão ligados entre si através de junções axonais que podem ser reorganizadas, o que permite a adaptação, o processamento e o

armazenamento de informação. De forma idêntica, as NN apresentam uma arquitetura idêntica de nódulos interconectados.

O *output* de cada nódulo será o *input* do próximo, criando uma matriz de camadas escondidas (*hidden layers*). Cada nódulo irá conferir um ganho próprio, que por sua vez irá permitir a amplificar ou enfraquecer o caminho da informação por esse nódulo.

Apesar de poder ser treinado para resolver problemas tanto de regressão como de classificação em dados cujas relações sejam complexas, as NN funcionam como caixas negras (*black-box*), que retiram interpretabilidade à forma como se chegou ao *output* final.

→ Random Forests (RF)

As florestas de decisão ou *random forests* são algoritmos preferencialmente de classificação, que geram resultados a partir de árvores de decisão (*decision trees*).

Cada uma destas árvores de decisão processa a informação através de nódulos desde a sua raiz (nódulo inicial) até aos seus ramos terminais (nódulo final) através de regras de decisão lógica. No final, se o problema for de classificação a maioria das classes nos nódulos terminais é apresentada como resultado, se o problema for de regressão, o resultado será a média.

O *output* é pouco provável que apresente *overfitting* quanto maior for o número de árvores de decisão na RF, no entanto este algoritmo tende a beneficiar variáveis de *input* com intervalos de valores mais elevados.

→ Extreme Gradient Boosting (XGboost)

Este algoritmo apresenta uma formatação idêntica à RF, no entanto submete-se a um processo no qual a cada árvore de decisão gerada é criada informação acerca dos seus erros residuais, informação essa que é transmitida à próxima árvore de decisão que terá estes erros residuais em conta na formação dos seus ramos, proporcionando assim um aumento da força preditiva da floresta.

As desvantagens deste algoritmo são idênticas às da RF, às quais se acrescenta um esforço computacional mais elevado.

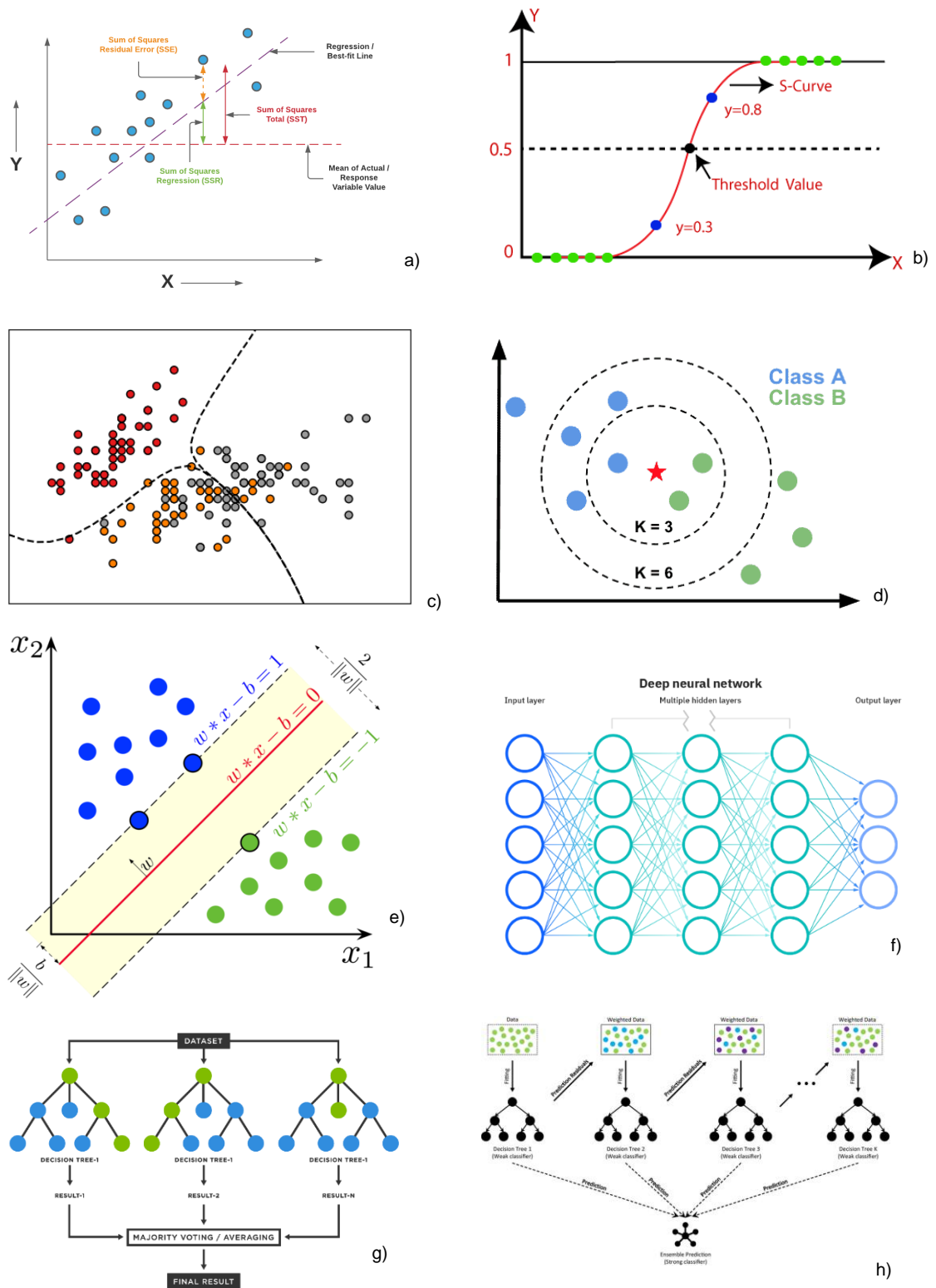


Figura 3.7 - Esquemática do funcionamento de cada modelo ML utilizado no presente estudo: a)  $LR^{25}$ ; b)  $LogR^{26}$ ; c)  $NB^{27}$ ; d)  $kNN^{28}$ ; e)  $SVM^{29}$ ; f)  $NN^{60}$ ; g)  $RF^{31}$ ; h)  $XGboost^{32}$ .

### 3.2.2. Seleção de variáveis e hiperparâmetros

A ordenação das variáveis para o treino dos algoritmos foi feita através da *Information Gain* (redução de entropia) para os modelos de classificação e por *RRelief* (avaliação da distância relativa entre valores previstos) para os modelos de regressão.

Os hiperparâmetros utilizados em cada um dos algoritmos pode ser observado na tabela 3.2.

Tabela 3-2 - Hiperparametros utilizados em cada um dos modelos.

<b>LR</b>	Ridge Regression (R2): $\alpha=0,11$
<b>LogR</b>	Lasso: L1 C=11
<b>kNN</b>	Number of neighbors: 5 Metric: Euclidean Weight: Uniform
<b>SVM</b>	Cost (C): 1 Regression loss epsilon ( $\epsilon$ ): 0,1 Kernel: polinomial
<b>NN</b>	Hidden layers: 100 Activation: ReLU Solver: Adam Regularization, $\alpha=40$ Maximal number interactions:200
<b>RF</b>	Number of trees: 10
<b>XGboost</b>	Number of trees: 100 Learning rate: 0,3 Lambda: 1

## 4. Materiais e Métodos

### 4.1. População associada à base de dados

Para a recolha de dados para esta dissertação utilizou-se a base de dados do projeto de investigação *Prediction of static lung volumes through spirometric parameters* (ID:13304738620) do Hospital da Luz de Lisboa (HLL) e da Learning Health.

Foram recolhidos retrospectivamente dados biológicos e antropométricos (sexo, idade, peso e altura), espirométricos (FEV<sub>1</sub> e FVC, antes e depois da prova de resposta ao BD, se realizada), pletismográficos (TLC e RV) e do estudo da transferência alvéolo-capilar de CO (TLCO), que foram introduzidos manualmente na plataforma Soarian Clinics®, pelos técnicos de Cardiopneumologia, aquando da realização dos exames. Posteriormente, foram calculados os parâmetros FEV<sub>1</sub>/FVC, RV/TLC, VC (pela subtração do RV à TLC) e todos os *z score* e percentagens do valor de referência. Dado que não estão disponíveis os dados da raça, todos os indivíduos foram identificados como caucasianos para cálculo do *z score* para a espirometria.

A base de dados foi recolhida e anonimizada pelo departamento de Informática do HLL, tendo sido garantida a confidencialidade dos participantes. O projeto foi aprovado pela Comissão de Ética para a Saúde do HLL.

No âmbito do referido projeto, as PFR foram realizadas em dois equipamentos *MasterScreen Body* da Jaeger® e um *Vyntus Body* da Vyaire®, todos com o software *SentrySuite 2.0*, sob as *guidelines* internacionais da ATS/ERS. As PCT foram realizadas em cabines de volume constante, duas delas com medição de débito através de pneumotacógrafo e estudo da transferência de CO com o hélio como *tracer gas*. No terceiro equipamento, a medição de débito foi realizada através de um sensor ultrassónico e a TLCO através do sistema *Rapid Gas Analysis*.

Para a prova de resposta ao BD foi administrada uma dose de 400 µg de salbutamol com utilização de câmara expansora. Em casos pontuais foi administrada uma dose de 160 µg de brometo de ipratrópio em substituição do salbutamol. Todos os parâmetros pré-prova de BD são considerados como basais.

## 4.2. Métodos de pré-processamento

A análise exploratória dos dados foi feita em linguagem *Python* utilizando o *Jupyter Notebook*. A aplicação e avaliação do desempenho dos modelos de ML foi feita através do *Orange Data Mining*.

Durante aproximadamente três anos e meio, entre janeiro de 2019 e julho de 2022, foram registadas na plataforma 14 646 PFR. Destas, foram selecionadas para projeto apenas as PFR que continham os dados antropométricos, espirométricos e pletismográficos completos, resultando num total de 8973 entradas. A amostra final, após remoção de *outliers* e outros erros de registo, apresentou um total de 8140 PFR.

Da amostra inicial de 14 646 PFR, existem 12 230 números de identificação única, correspondentes a indivíduos que realizaram apenas uma vez as PFR neste período. Assim, é razoável assumir que possam existir diferentes PFR referentes ao mesmo participante.

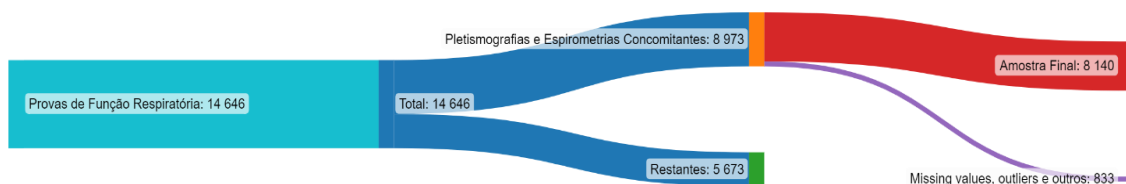


Figura 4.1 - Diagrama de Sankey da depuração da base de dados até à amostra final.

O primeiro passo após a aquisição dos dados foi o tratamento dos mesmos. Neste sentido, entradas com claros *outliers* (i.e., altura de 769 cm) foram eliminadas. A identificação destes *outliers* foi realizada através dos limites de cada parâmetros e/ou visualização gráfica (figura 4.2).

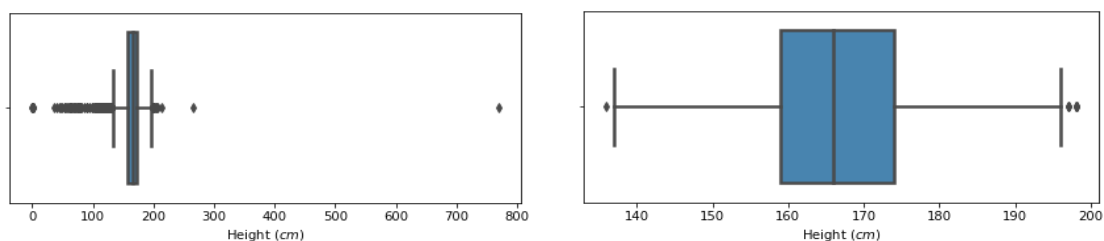


Figura 4.2 - *Boxplot* da distribuição da altura, antes e após a remoção de *outliers* severos.

Para tal foram criados critérios de inclusão, após a identificação de *outliers*:

- Idade entre os 18 e os 80 anos;
- Altura entre os 130 e os 200 cm;
- Peso entre os 30 e os 200 kg;
- TLC inferior a 15 L;
- RV inferior a 10 L;
- RV/TLC inferior a 0,90;
- FEV<sub>1</sub>/FVC entre 0,10 e 1,00;
- FEV<sub>1</sub>/FVC *z-score* inferior a 7.

Entradas com valores nulos (*missing values*) só foram verificadas em parâmetros após prova de broncodilatação e TLCO, devido à natureza da seleção da amostra (todas PFR com *missing values* nos dados antropométricos, espirométricos ou pletismográficos foram eliminadas).

Neste projeto foram calculados, posteriormente à aquisição dos dados, os valores previstos e respetivos *z score* através das equações de referência para a espirometria<sup>17</sup> dos 3 aos 95 anos, para a TLCO<sup>15</sup> em caucasianos dos 5 aos 85 anos e para os volumes pulmonares estáticos<sup>12</sup> para indivíduos de ascendência europeia dos 5 aos 80 anos.

### 4.3. Estratégias para a análise exploratória dos dados

A exploração dos dados de forma a maximizar a capacidade preditiva dos modelos ML foi realizada pela seguinte ordem:

- i. Aquisição e limpeza da base de dados

Os dados tabulares fornecidos pelo HLL foram recolhidos diretamente de registos feitos por profissionais de saúde no âmbito do seu trabalho, pelo que continham algumas inconformidades que tiveram de ser ultrapassadas.

Das variáveis iniciais, foram utilizadas as seguintes: idade, sexo, peso, altura, TLC, RV, TLCO e FEV<sub>1</sub> e FVC pré e após prova de resposta ao BD. Todas as restantes foram eliminadas. Seguidamente, com recurso à calculadora *online GLI Lung Function Calculator*<sup>33</sup>, calculou-se as percentagens do valor previsto e os *z score* para TLC, RV,

RV/TLC, TLCO, FEV<sub>1</sub>, FVC e FEV<sub>1</sub>/FVC. Todas as instâncias com valores nulos presentes em qualquer uma das seguintes variáveis foram removidos: idade, sexo, peso, altura, TLC, RV e FEV<sub>1</sub> e FVC pré prova de resposta ao BD.

Após a limpeza e a introdução dos *z score*, foi possível criar variáveis como: IMC, IMC categórico, resposta ao BD categórica, gravidade categórica (criada a partir do FEV<sub>1</sub> *z score*), categorias resultantes da interpretação da espirometria, da PCT e da TLCO.

#### ii. Análise univariada

Para cada variável numérica foram avaliados parâmetros estatísticos básicos como a média, o desvio padrão, o valor mínimo, o valor máximo, a mediana e os interquartis 25 e 75. Identificou-se a presença de *outliers* severos na visualização de gráficos *boxplot*, que foram de seguida removidos, estando na origem dos critérios de inclusão do presente estudo. Avaliou-se nesta fase a assimetria e o achatamento de cada uma das variáveis com o auxílio visual de histogramas.

#### iii. Análise bivariada

Nesta fase de forma a identificar as relações entre pares de variáveis foram utilizados gráficos bidimensionais: *scatter plots* e *line plots*. O cálculo de coeficientes de Pearson quantificou estas relações.

#### iv. Análise multivariada

Para avaliar as ligações entre grupos de variáveis, foram criados *heat maps* e tabelas de dupla entrada.

### 4.4. Parâmetros de avaliação de modelos

Foram consideradas diferentes métricas para avaliar o desempenho dos modelos, de acordo com o alvo em questão.

Os **modelos de classificação** foram avaliados com base na sensibilidade, valor preditivo positivo, precisão, área abaixo da curva *Receiver Operating Characteristic* (ROC) e *F<sub>1</sub> score*.<sup>20,21</sup>

A tabela de contingência ou *confusion matrix* é um instrumento útil para quantificar o desempenho de um modelo de classificação, e da qual é possível retirar as restantes métricas para estudo do modelo. Os resultados são categorizados em verdadeiros positivos (quando o modelo prevê corretamente um valor positivo real), verdadeiros negativos (quando o modelo prevê corretamente um valor negativo real), falsos positivos/erro tipo I (quando o modelo prevê incorretamente um valor negativo real) e falsos negativos/erro tipo II (quando o modelo prevê incorretamente um valor positivo real). O modelo ideal apresenta as células preenchidas na diagonal e as restantes células com o valor 0. <sup>20,21</sup>

		Classificação Prevista		
		Positivo (+)	Negativo (-)	
Classificação Real	Positivo (+)	Verdadeiro Positivo (VP)	Falso Negativo (FN) Erro tipo II	<b>Sensibilidade</b> ( <i>recall</i> ) $\frac{VP}{VP + FN}$
	Negativo (-)	Falso Positivo (FP) Erro tipo I	Verdadeiro Negativo (VN)	<b>Especificidade</b> $\frac{VN}{(VN + FP)}$
		<b>Valor Preditivo Positivo</b> ( <i>precision</i> ) $\frac{VP}{VP + FP}$	<b>Valor Preditivo Negativo</b> $\frac{VN}{VN + FN}$	<b>Precisão</b> ( <i>accuracy</i> ) $\frac{VP + VN}{VP + FP + VN + FN}$

Figura 4.3 - Esquema da tabela de contingência, com o significado de cada quadrante e dos cálculos associados para avaliação dos modelos de classificação.

A sensibilidade ou *recall* indica a proporção de resultados verdadeiros positivos de todos os resultados que o modelo previu corretamente, ou seja, quantifica o quão bem o classificador é capaz de detetar casos de alteração ventilatória na população com patologia.<sup>20,21</sup>

O valor preditivo positivo ou *precision* indica a proporção de resultados verdadeiros de todos os resultados que o modelo previu como positivos, ou seja, sugere o quão confiável é o algoritmo se o resultado for positivo (neste caso, com dada alteração ventilatória).<sup>20,21</sup>

A precisão ou *accuracy* indica a proporção de resultados que o modelo previu corretamente (positivos e negativos) de entre todos os resultados.<sup>20,21</sup>

O  $F_1$  score é a média harmónica entre a sensibilidade e o valor preditivo positivo, o que torna este parâmetro mais robusto do que a precisão.<sup>20,21</sup>

$$F_1 \text{ score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

A área abaixo da curva ROC (AUC) usa a relação entre o valor preditivo positivo e o valor preditivo negativo de forma a avaliar o desempenho de um modelo. Quanto maior for o valor da AUC, melhor o desempenho do modelo. Se a AUC for igual a 0,5, o modelo não é melhor a prever um resultado do que um classificador aleatório. Se a AUC for 1, o modelo é perfeito e identifica sem qualquer erro os valores reais.<sup>20,21</sup>

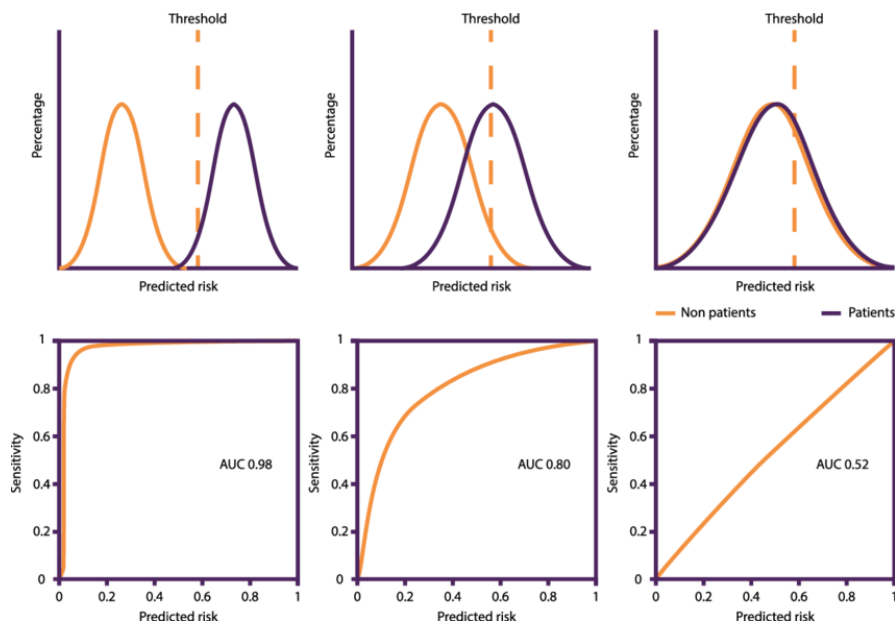


Figura 4.4 - Distribuição de probabilidades e AUC associada.<sup>22</sup>

Os **modelos de regressão** foram avaliados com base na Erro Absoluto Médio ou *Mean Absolut Error* (MAE), Erro Quadrático Médio ou *Mean Square Error* (MSE), Raiz do Erro Quadrático Médio ou *Root of Mean Square Error* (RMSE) e coeficiente de determinação ( $R^2$ ).<sup>20,21</sup>

O erro, na avaliação entre de modelos de regressão, é-nos dado pela seguinte fórmula:<sup>20,21</sup>

$$e = yi - \hat{y}i$$

em que  $yi$  é o valor real e  $\hat{y}i$  é o valor previsto pelo modelo.

Várias métricas podem ser utilizadas para avaliar o quão afastado está o modelo da realidade, entre as quais o MAE, que é o módulo da diferença média entre o valor real e o previsto. Esta métrica apresenta uma vantagem em relação às restantes pois apresenta as mesmas unidades de medida dos valores, que acrescenta interpretabilidade à leitura. No entanto, não indica se o erro está a sobrevalorizar ou a subvalorizar em relação ao real, nem penaliza erros graves.<sup>20,21</sup>

$$MAE = \frac{1}{N} \sum_{i=1}^N |yi - \hat{y}i|$$

O MSE, tal como o MAE, calcula a diferença média entre o valor real e o valor previsto, contudo, essa diferença é elevada ao quadrado. Isto permite penalizar os erros maiores em relação ao valor real, mas perde interpretabilidade pois os resultados obtidos são o quadrado das unidades de medida.<sup>20,21</sup>

$$MSE = \frac{1}{N} \sum_{i=1}^N (yi - \hat{y}i)^2$$

De forma a colmatar esta situação, surge o RMSE. O RMSE apresenta valores mais próximos das unidades de medida mantendo a penalização de erros grandes, estando assim também mais suscetível a *outliers*.<sup>20,21</sup>

$$RMSE = \frac{1}{N} \sqrt{\sum_{i=1}^N (yi - \hat{y}i)^2}$$

Outra métrica bastante utilizada na avaliação de modelos preditivos de regressão é o  $R^2$ . O  $R^2$  é muito útil pois os seus valores variam entre 0 e 1 (boa interpretabilidade) e indica a capacidade do modelo de explicar a variância da amostra.

20,21

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2}$$

A qualidade de cada algoritmo deverá ser avaliada de acordo com o problema em questão. Em algumas situações, uma AUC de 0,9 pode não ser suficientemente elevada para que se considere um bom modelo, em outras situações uma AUC de 0,7 poderá ser excelente. O mesmo se aplica aos problemas de regressão com o  $R^2$ .

Neste sentido, definiu-se uma classificação qualitativa para os modelos, em que uma AUC (ou  $R^2$ ) entre 0,9 e 1 o modelo teve um desempenho excelente, entre 0,8 e 0,9 o desempenho foi bom, entre 0,7 e 0,8 o desempenho foi aceitável e abaixo de 0,7 o desempenho será considerado fraco.

## 5. Resultados

### 5.1. Análise exploratória dos dados

Dos 8140 indivíduos que compunham a amostra final, 4556 eram do sexo feminino (55,97%) e 3584 do sexo masculino (44,03%). A idade média foi de  $53 \pm 16$  anos. A altura média foi de  $166 \pm 10$  cm e o peso médio foi de  $75,4 \pm 17,3$  kg.

Os dados estatísticos dos parâmetros das PFR estudados estão representados na tabela 5.1.

Tabela 5-1 - Estatística descritiva univariada das variáveis obtidas das PFR.

	FEV <sub>1</sub> (L)	FVC (L)	FEV <sub>1</sub> /FVC	TLC <sub>CO</sub> (mmol/min/kPa)	TLC (L)	RV (L)	RV/TLC	VC (L)
<b>n</b>	8140	8140	8140	6429	8140	8140	8140	8140
<b>média</b>	2,85	3,72	0,76	6,60	5,93	2,17	0,37	3,76
<b>desvio-padrão</b>	0,95	1,11	0,10	2,14	1,29	0,71	0,10	1,11
<b>mínimo</b>	0,37	0,65	0,11	0,80	1,54	0,34	0,06	0,65
<b>quartil 25</b>	2,17	2,92	0,72	5,10	4,99	1,70	0,30	2,95
<b>quartil 50</b>	2,79	3,61	0,78	6,39	5,78	2,10	0,36	3,64
<b>quartil 75</b>	3,46	4,43	0,83	7,93	6,81	2,53	0,43	4,46
<b>máximo</b>	6,16	8,21	0,99	15,00	11,39	6,91	0,84	9,08

Na avaliação de cada parâmetro individualmente, em termos de assimetria (*skewness*), a maioria dos parâmetros das PFR apresentou valores entre -0,5 e 0,5, o que demonstra uma simetria na distribuição dos dados: 0,38 para a TLC, 0,35 para o FEV<sub>1</sub>, 0,42 para a FVC e 0,45 para o RV/TLC. No entanto, o FEV<sub>1</sub>/FVC apresentou uma assimetria positiva significativa (-1,26) e o RV uma assimetria negativa significativa (1,03), ou seja, estes valores indicam a possibilidade de *outliers* de valores reduzidos (no caso do FEV<sub>1</sub>/FVC) ou elevados (no caso do RV) em relação à normalidade.

Na análise dos parâmetros antropométricos, apenas o peso (0,77) apresentou uma distribuição assimétrica.

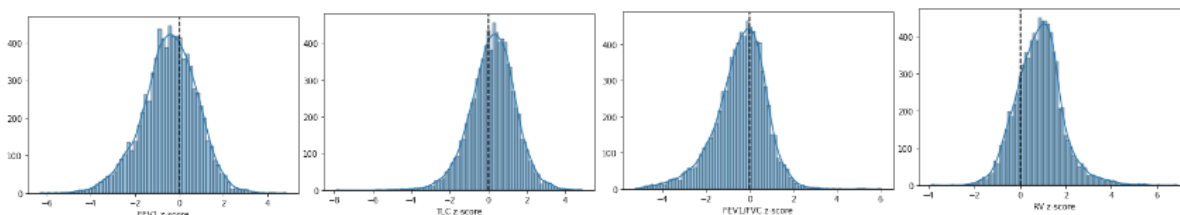


Figura 5.1 – Histogramas dos z score de FEV<sub>1</sub>, TLC, FEV<sub>1</sub>/FVC e RV com média (z score=0) evidenciada.

Na medida estatística de achatamento (*kurtosis*), a maioria dos parâmetros apresentou uma distribuição mesocúrtica, ou seja, idêntica à distribuição normal.

A idade (-0,85) foi o parâmetro relevante com um achatamento superior à normalidade (platicúrtica), enquanto o peso, o FEV<sub>1</sub>/FVC e o RV apresentam um achatamento inferior à normalidade (leptocúrtica).

Em termos de análise estatística bivariada, uma característica comum à maioria dos volumes pulmonares foi a sua variação ao longo da idade de um indivíduo. Conforme se observa na figura 5.2, a FVC e o FEV<sub>1</sub> começaram a decair por volta dos 35 anos, mas o FEV<sub>1</sub> decresceu a um ritmo ligeiramente mais acelerado, traduzindo-se numa diminuição da relação FEV<sub>1</sub>/FVC com o passar dos anos.

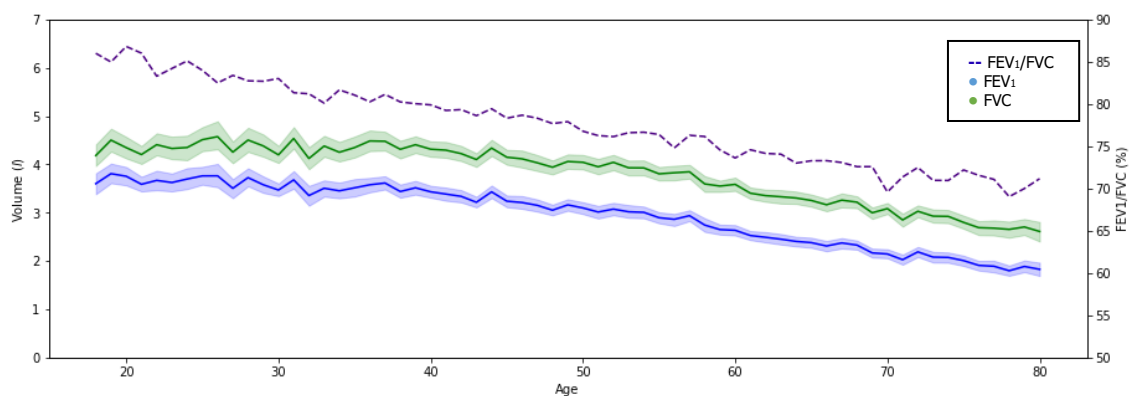


Figura 5.2 - Progresso do FEV<sub>1</sub>/FVC, FEV<sub>1</sub> e FVC ao longo da idade.

No que respeita aos parâmetros pletismográficos avaliados, a TLC manteve-se relativamente constante (com ligeira diminuição), enquanto o RV aumenta gradualmente a partir dos 35 anos. Atendendo a que  $TLC=VC+RV$ , aos 20 anos de idade o RV é, em média, 33% de TLC enquanto a VC representa cerca de 66%. Aos 80 anos a percentagem do RV e da VC, em relação à TLC, é aproximadamente 50%, como é possível observar na figura 5.3.

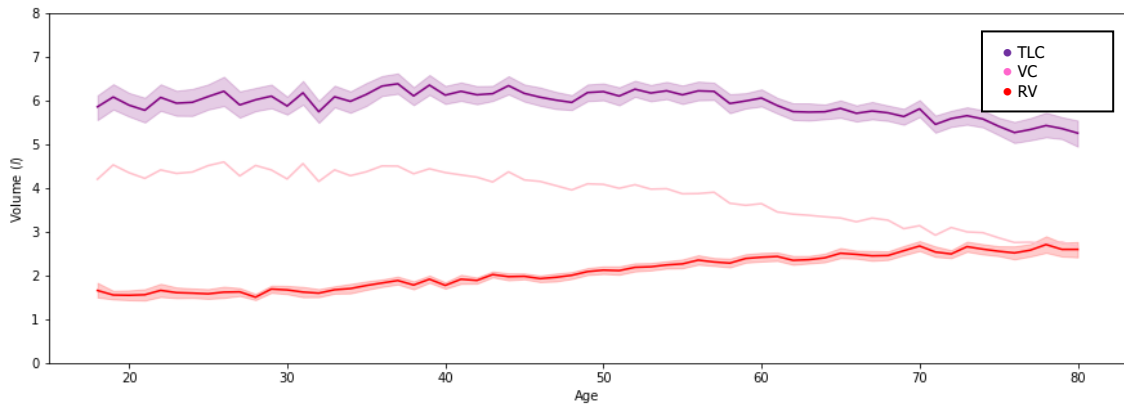


Figura 5.3 - Progresso da TLC, VC e RV ao longo da idade.

Analisando o RV e o  $FEV_1$ , verifica-se que apresentam um comportamento inverso, com o  $FEV_1$  a decair gradualmente enquanto o RV aumenta conforme a idade dos indivíduos aumenta (figura 5.4).

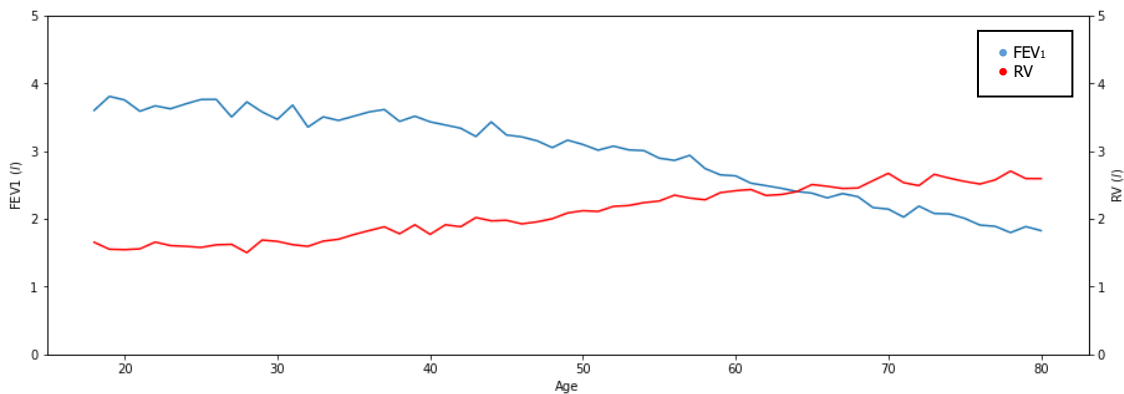


Figura 5.4 - Progresso do  $FEV_1$  e do RV ao longo da idade.

Substituindo o RV pela relação  $RV/TLC$ , e retirando a variável idade, é possível observar que a relação entre  $FEV_1$  e  $RV/TLC$  exibe um comportamento menos linear e mais polinomial, como é possível verificar na figura 5.5a. Em todas as categorias de interpretação da espirometria, os indivíduos com o  $RV/TLC$  aumentado eram aqueles com menores valores de  $FEV_1$ , associando assim o aumento do  $RV/TLC$  com a gravidade da alteração ventilatória (figura 5.5b).

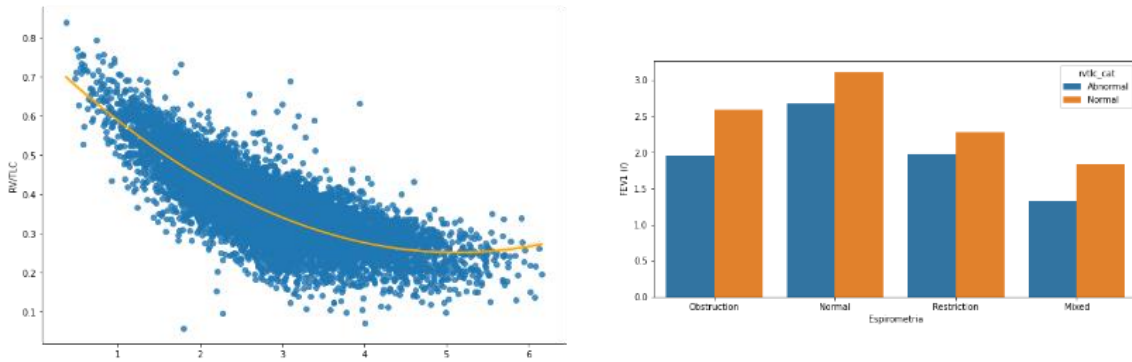


Figura 5.5 - Relação entre RV/TLC e FEV1 (a) e distribuição do FEV1 pelas categorias da espirometria, divididas pela alteração de RV/TLC (b).

Os indivíduos do sexo masculino apresentaram, em média, volumes pulmonares mais elevados do que os indivíduos do sexo feminino, como pode ser observado na figura 5.6a. Estas diferenças esbatem-se, como era expectável, quando se avaliam os *z score* de cada parâmetro, divididos pelo sexo (figura 5.6b).

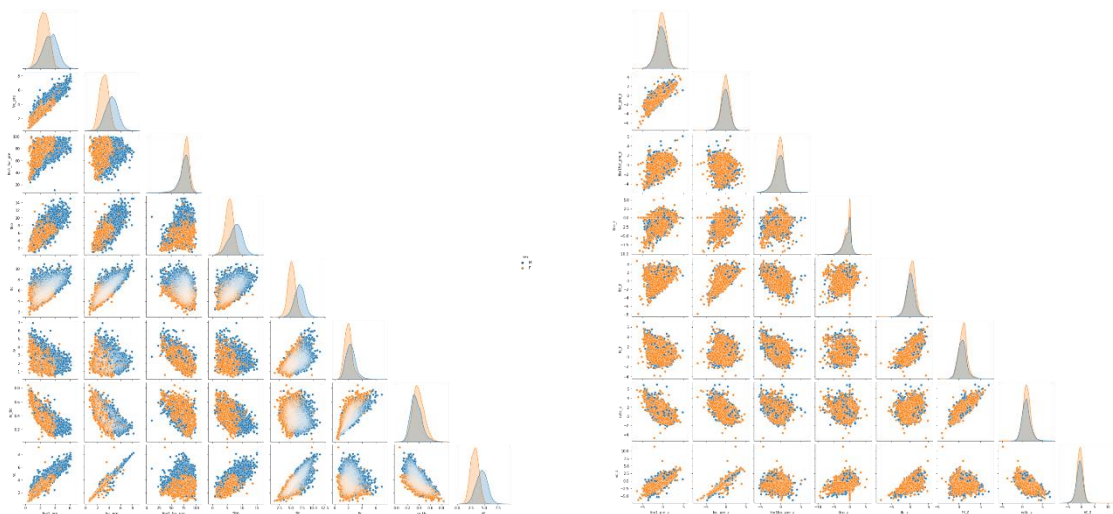


Figura 5.6 – Distribuição da correlação dos valores absolutos (a) e dos *z score* (b) por sexo.

Estas diferenças de volume entre sexos são menores conforme se avança no espectro da gravidade da alteração ventilatória, com indivíduos do sexo feminino e do

sexo masculino com valores absolutos de FEV<sub>1</sub> idênticos na presença de alteração ventilatória grave (figura 5.7).

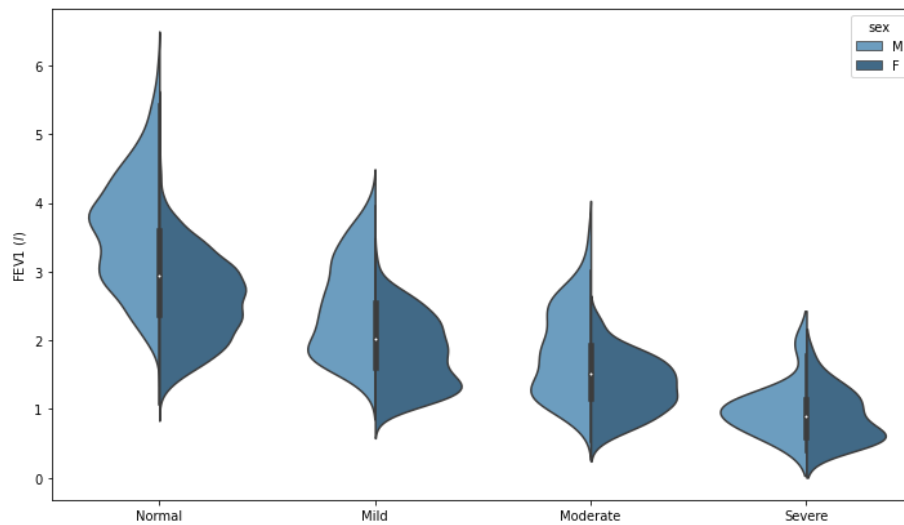


Figura 5.7 - Violin plots das categorias da espirometria divididas por sexo, em que se observa uma diminuição das diferenças entre sexos com o agravamento da alteração ventilatória.

Com o aumento do peso não se verificou um aumento evidente do valor absoluto dos parâmetros associados ao *air trapping*, como é possível observar na figura 5.8.

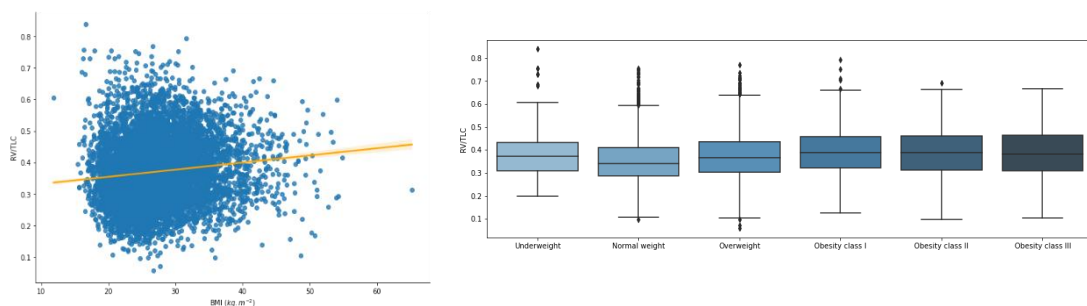


Figura 5.8 - Relação entre o RV/TLC e o IMC (a) e *box plots* da distribuição de RV/TLC pelas categorias de IMC (b).

Ainda assim, quando se estudou o quanto os valores da função respiratória se afastam do normal (*z score*), verificou-se que, como consequência do aumento do IMC em relação ao peso ideal, existiu um aumento gradual dos *z score* de RV/TLC, e um decréscimo dos *z score* do FEV<sub>1</sub>, com exceção os indivíduos com baixo peso, que apresentam em média os *z score* do FEV<sub>1</sub> diminuídos e os *z score* do RV/TLC aumentados.

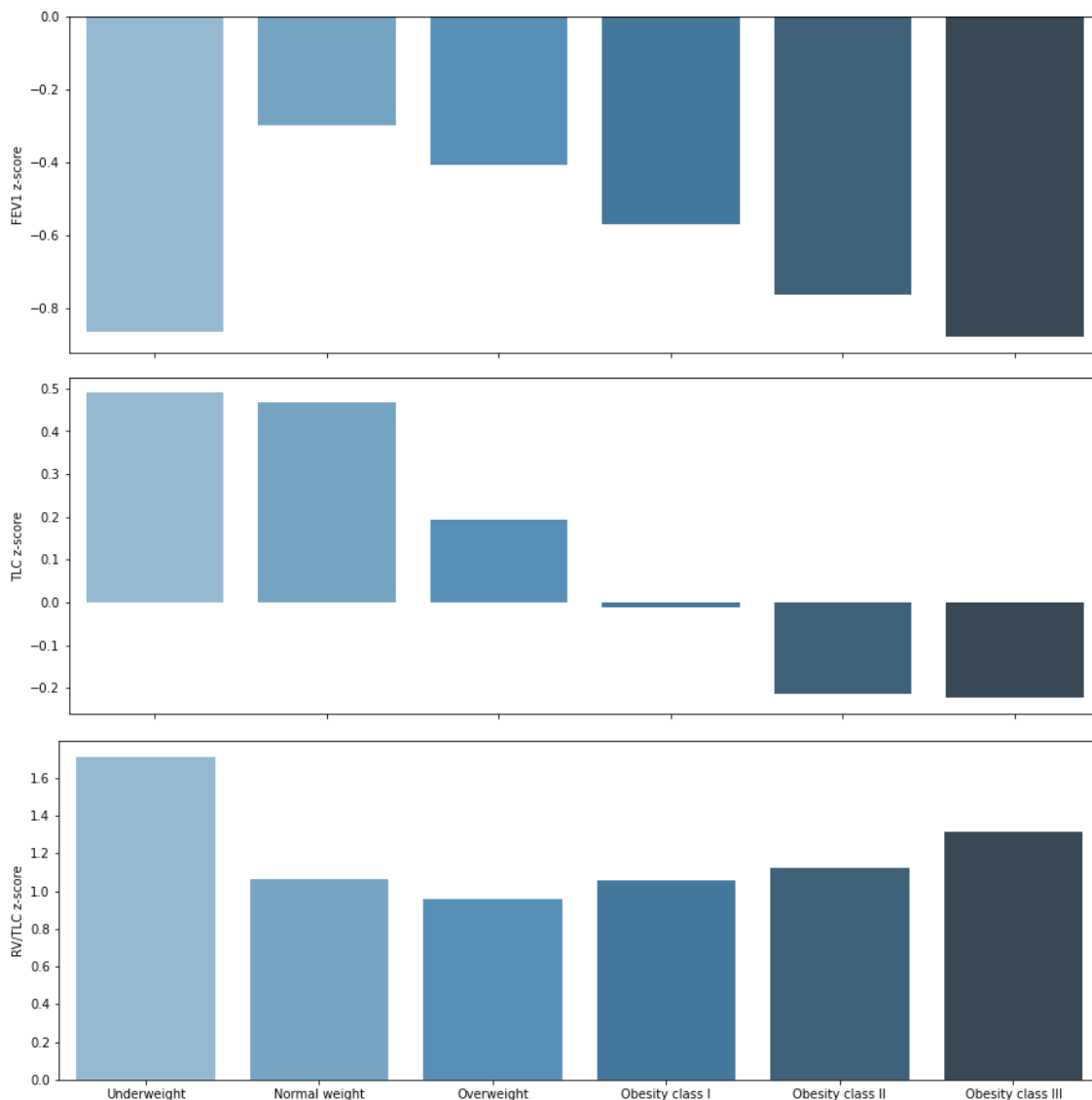


Figura 5.9 - Valores médios de *z score* do FEV<sub>1</sub>, TLC e RV/TLC de acordo com as categorias de IMC.

O *z score* da TLC decresceu progressivamente com as classes de IMC mais elevadas.

O FEV<sub>1</sub> apresentou um coeficiente de correlação de Pearson (R) de -0,78 com RV/TLC. Esta correlação foi mais forte do que cada um dos componentes do RV/TLC separadamente (RV e TLC). A FVC aparenta ter um elevado potencial para a identificação de variações da TLC, com um R de 0,83.

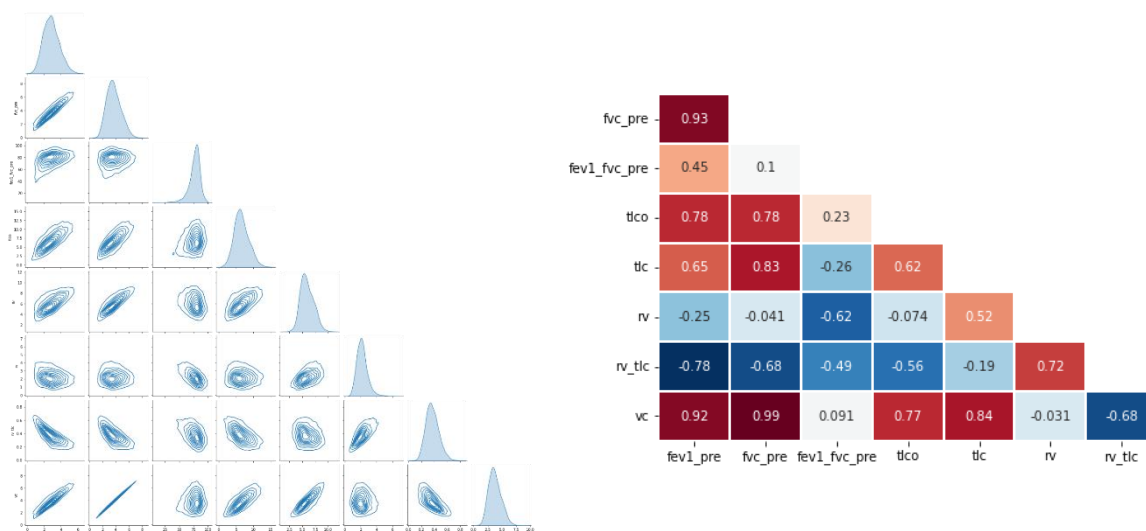


Figura 5.10 – Gráficos bidimensionais das variáveis das PFR (a) e *heat map* com as correlações de Pearson (b).

Em termos de categorização espirométrica, uma espirometria sem alterações foi o padrão mais vezes registado, seguido da alteração obstrutiva, restritiva e mista, como é possível observar na tabela 5.2.

Tabela 5-2 - Distribuição da prevalência das alterações observadas na espirometria.

<b>Alteração espirométrica</b>	<i>n</i>	%
Normal	6428	79,0
Obstrutiva	980	12,0
Restritiva	584	7,2
Mista	148	1,8

Em termos de categorização da PCT, a distribuição normal de volumes estáticos foi o padrão mais vezes registado, seguido de *air trapping*, como é possível observar na tabela 5.3.

Tabela 5-3 - Distribuição da prevalência das alterações observadas na PCT.

<b>Alteração pletismográfica</b>	<i>n</i>	%
Volumes estáticos normais	5376	66,0
<i>Air trapping</i>	1504	18,5
Volumes estáticos proporcionalmente aumentados	431	5,3
Hiperinsuflação pulmonar	359	4,4
Restrição simples	338	4,2
Restrição complexa	108	1,3
Mista	24	0,3

De salientar que o RV/TLC acima do ULN que está associado ao *air trapping*, à hiperinsuflação, à restrição complexa e à alteração mista, esteve presente em 24,4% dos indivíduos da amostra, sendo o elemento da PCT mais vezes alterado.

Relativamente à gravidade da alteração ventilatória, avaliada pelo valor do *z score* do FEV<sub>1</sub> pré-prova de broncodilatação, 6875 (84,5%) indivíduos apresentaram um valor normal, 789 (9,7%) ligeiramente diminuído, 434 (5,3%) moderadamente diminuído e 42 (0,5%) gravemente diminuído.

Tabela 5-4 – Tabela cruzada entre as categorias interpretativas da espirometria e da PCT.

<b>Alteração pletismográfica</b>	<b>Alteração espirométrica</b>	<b>Normal</b>	<b>Obstrução</b>	<b>Restrição</b>	<b>Misto</b>
Volumes estáticos normais		4917	401	54	4
<i>Air trapping</i>		897	272	232	103
Volumes estáticos proporcionalmente aumentados		343	88	0	0
Hiperinsuflação pulmonar		123	215	0	21
Restrição simples		147	0	191	0
Restrição complexa		1	0	0	107
Mista		0	4	0	20

Relativamente à resposta ao fármaco broncodilatador, cerca de 11,8% dos indivíduos a quem foi realizada prova de broncodilatação responderam positivamente, de acordo com os critérios mais recentes.<sup>5</sup> Do total de alterações obstrutivas avaliadas por espirometria, 23,8% responderam positivamente à prova de broncodilatação (enquanto 6% dos indivíduos com espirometria normal tiveram um resultado idêntico).

Em termos de caracterização do estado da transferência alvéolo-capilar da amostra (acessória para os objetivos do estudo), observou-se que uma TLCO anormalmente diminuída surgiu com prevalência mais elevada em alterações ventilatórias cujo impacto estrutural é maior, ou seja, nas quais a TLC sofre maiores alterações patológicas (restrição complexa, restrição simples, alteração mista e hiperinsuflação pulmonar). De acordo com a tabela 5.5, observa-se ainda que valores anormalmente elevados de TLC, mas com a relação com RV normal (volumes estáticos proporcionalmente aumentados), são os que apresentam uma prevalência menor de alterações da transferência alvéolo-capilar, o que indica que esta categoria da PCT pode não ter um significado patológico e apenas traduzir uma característica estrutural do pulmão desses indivíduos.

Tabela 5-5 - Tabela cruzada entre as categorias interpretativas da espirometria e da TLCO, com a prevalência da TLCO diminuída por categoria da PCT.

<b>Alteração pletismográfica</b>	<b>TLCO</b>	Diminuído	Normal	% Diminuído
Volumes estáticos normais		1004	4372	18,7
<i>Air trapping</i>		401	1103	26,7
Hiperinsuflação pulmonar		128	231	35,7
Volumes estáticos proporcionalmente aumentados		49	382	11,4
Restrição simples		190	148	56,2
Restrição complexa		66	42	61,1
Mista		11	13	45,8

## 5.2. Modelos

Dada a prevalência relativamente elevada de *air trapping* (1504 em 8140, 18,5% dos exames realizados) que não pode ser medida diretamente por espirometria, estudaram-se diferentes modelos de ML de forma a identificar indivíduos com um RV/TLC aumentado (*z score* superior a 1,645).

De um total de 26 variáveis espirométricas e antropométricas, foram selecionadas as 20 com maior ganho de informação para os modelos, sendo que destas, as cinco com maior ganho de informação foram (por ordem decrescente): *z score* FEV<sub>1</sub> basal; percentagem do valor previsto do FEV<sub>1</sub>; gravidade da alteração ventilatória; *z score* FVC basal e percentagem do valor previsto da FVC.

Na tabela 5.6 estão expostos os resultados dos modelos testados com a técnica de reamostragem *cross validation 20-fold* (por ordem decrescente de F<sub>1</sub> score).

Tabela 5-6 - Resultados do desempenho dos modelos utilizados na classificação dicotômica de RV/TLC.

Modelos ML	AUC	Precisão	F <sub>1</sub> score	Sensibilidade (precision)	Valor preditivo positivo (recall)
<b>XGradient Boosting</b>	0,874	0,834	0,824	0,835	0,834
<b>Random Forest</b>	0,841	0,826	0,814	0,818	0,821
<b>kNN</b>	0,808	0,814	0,805	0,803	0,814
<b>Neural Network</b>	0,820	0,816	0,787	0,812	0,816
<b>Logistic Regression</b>	0,846	0,768	0,780	0,811	0,768
<b>Naive Bayes</b>	0,824	0,744	0,759	0,795	0,744
<b>SVM</b>	0,588	0,764	0,676	0,758	0,764

Todos os modelos apresentaram resultados idênticos, no entanto com diferentes prevalências de erros tipo I e tipo II, como é possível verificar na figura 5.11.

		Predicted		
		Abnormal	Normal	Σ
Actual	Abnormal	1027	962	1989
	Normal	386	5765	6151
Σ		1413	6727	8140

		Predicted		
		Abnormal	Normal	Σ
Actual	Abnormal	993	996	1989
	Normal	514	5637	6151
Σ		1507	6633	8140

		Predicted		
		Abnormal	Normal	Σ
Actual	Abnormal	1494	495	1989
	Normal	1396	4755	6151
Σ		2890	5250	8140

Figura 5.11 - Tabelas de contingência do XGboost (a), kNN (b) e LogR (c).

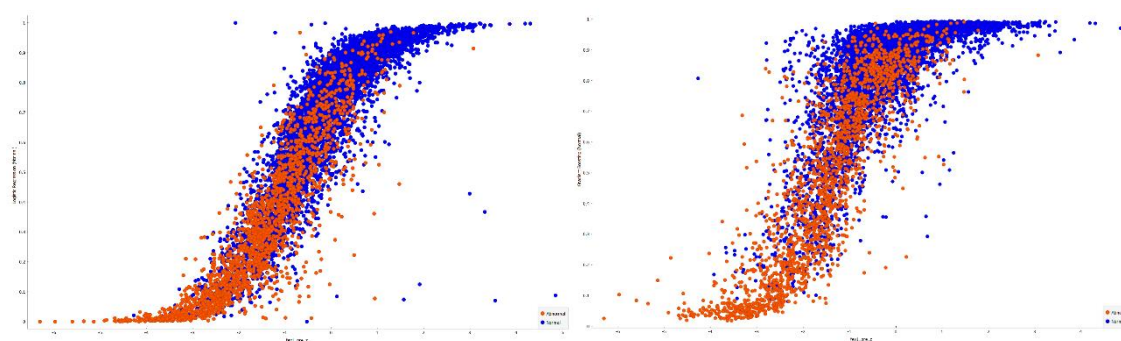


Figura 5.12 - Distribuição de probabilidades da LogR (a) e XGboost (b) para RV/TLC normal consoante o z score do FEV<sub>1</sub> com XGboost a apresentar uma maior precisão na identificação de um RV/TLC normal.

De seguida, procedeu-se a uma técnica de empilhamento (*stacking*) de diferentes modelos, especificamente XGboost, RF, kNN e LogR, e obtiveram-se os seguintes resultados:

Tabela 5-7 - Resultados do desempenho do empilhamento de modelos na classificação dicotômica de RV/TLC.

Modelos	AUC	Precisão	F <sub>1</sub> score	Precision	Recall
<b>Stack</b>	0,861	0,786	0,796	0,817	0,786

		Predicted		Σ
		Abnormal	Normal	
Actual	Abnormal	1473	516	1989
	Normal	1060	5091	6151
Σ		2533	5607	8140

Figura 5.13 - Tabela de contingência do modelo de empilhamento.

Com o intuito de prever qual o valor do z score da relação RV/TLC, utilizou-se a mesma técnica amostral com 20 variáveis em que as cinco primeiras com maiores ganhos de informação foram: resposta categórica ao BD, FEV<sub>1</sub> após prova de BD, FVC após prova de BD, FEV<sub>1</sub>/FVC após prova de BD e idade. A tabela 5.8 apresenta os resultados obtidos por ordem crescente de MAE.

Tabela 5-8 - Resultados do desempenho dos modelos utilizados na regressão do z score de RV/TLC.

Modelos ML	MSE	RMSE	MAE	R <sup>2</sup>
<b>Linear Regression</b>	0,582	0,763	0,595	0,442
<b>Neural Network</b>	0,641	0,801	0,622	0,386
<b>kNN</b>	0.642	0.801	0,625	0,386

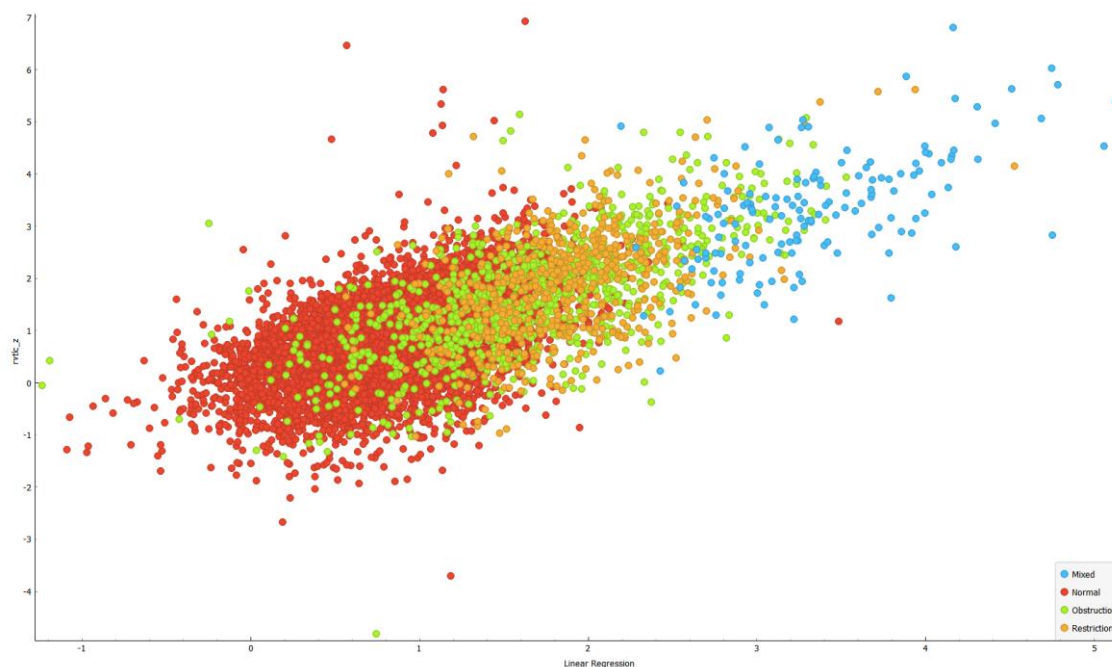


Figura 5.14 – Scatter plot entre o valor real e o valor previsto pela LR do z score de RV/TLC, colorida pelas categorias da espirometria, demonstrando um subvalorização das previsões.

Utilizando a mesma metodologia, tentou prever-se o valor absoluto de RV/TLC, com os seguintes resultados:

Tabela 5-9 - Resultados do desempenho dos modelos utilizados na regressão do valor absoluto de RV/TLC.

<b>Modelos ML</b>	<b>MSE</b>	<b>RMSE</b>	<b>MAE</b>	<b>R<sup>2</sup></b>
<b>Linear Regression</b>	0,003	0,051	0,040	0,752
<b>kNN</b>	0,003	0,056	0,044	0,700
<b>Neural Network</b>	0,007	0,087	0,069	0,260

Neste modelo, das 20 variáveis, as cinco com maior peso por ordem decrescente foram: resposta categórica ao BD; FEV<sub>1</sub> após prova de BD; FVC após prova de BD; FEV<sub>1</sub>/FVC após prova de BD e IMC categórico.

Aplicando os mesmos métodos, foi abordado o problema de regressão com o valor absoluto de RV como alvo. As cinco variáveis mais informativas foram as mesmas para a previsão do RV/TLC com os seguintes resultados:

Tabela 5-10 - Resultados do desempenho dos modelos utilizados na regressão do valor absoluto de RV.

<b>Modelos ML</b>	<b>MSE</b>	<b>RMSE</b>	<b>MAE</b>	<b>R<sup>2</sup></b>
<b>Linear Regression</b>	0,230	0,479	0,372	0,539
<b>kNN</b>	0,258	0,508	0,378	0,482
<b>Neural Network</b>	0,262	0,512	0,395	0,473

De seguida, fez-se a previsão do valor absoluto da TLC, essencial para a identificação de alterações ventilatórias restritivas e mistas (tabela 5.11).

Das 20 variáveis utilizadas, as cinco com mais peso para os modelos foram, por ordem decrescente, FEV<sub>1</sub> após prova de BD; FVC após prova de BD; FEV<sub>1</sub>/FVC após prova de BD; IMC categórico e idade.

Tabela 5-11 - Resultados do desempenho dos modelos utilizados na regressão do valor absoluto da TLC.

Modelos ML	MSE	RMSE	MAE	R <sup>2</sup>
Linear Regression	0,241	0,491	0,380	0,856
Neural Network	0,277	0,527	0,409	0,834
kNN	0,368	0,606	0,471	0,781

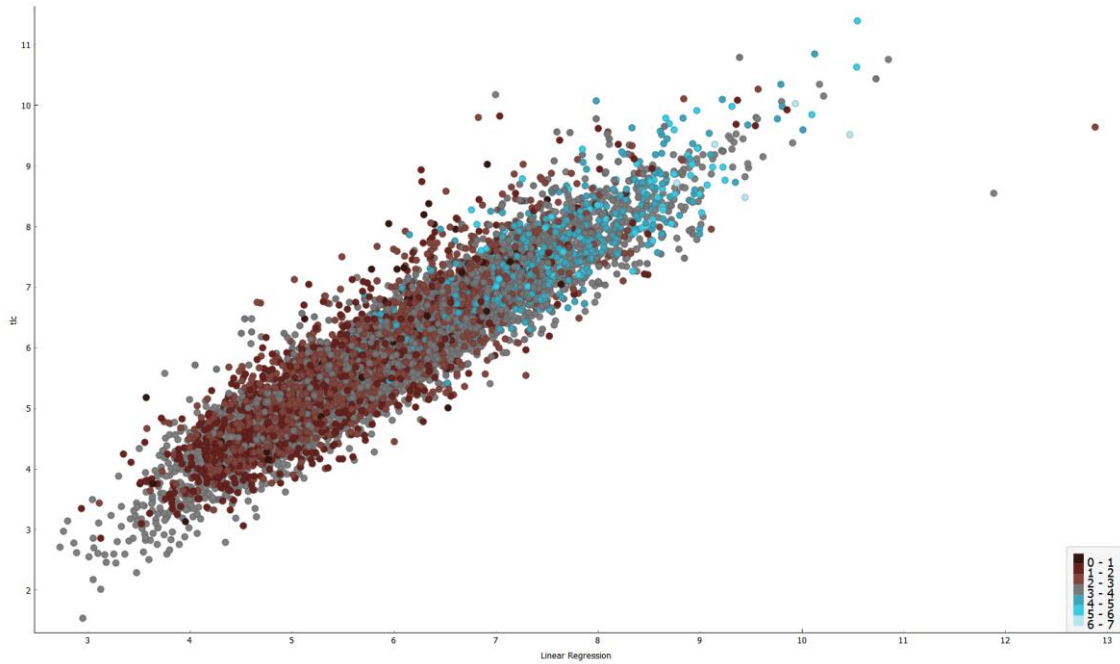


Figura 5.15 – Scatter plot entre valor real e o valor previsto pela LR do z score de TLC, colorido de acordo com o FEV<sub>1</sub> pós BD.

Por último, aplicou-se os modelos preditivos para a classificação categórica das alterações pletismográficas. Neste sentido, as cinco variáveis com maior peso foram: FVC basal; z score FVC basal; z score FEV<sub>1</sub> basal; percentagem do valor previsto do FEV<sub>1</sub> e categoria de alteração espirométrica.

		Predicted							$\Sigma$
		Air trapping	Complex Restriction	Hyperinflation	Large lung	Mixed	Normal static volumes	Simple Restriction	
Actual	Air trapping	683	15	46	1	8	689	62	1504
	Complex Restriction	20	56	0	0	1	0	31	108
	Hyperinflation	74	0	97	18	3	166	1	359
	Large lung	0	0	10	168	0	253	0	431
	Mixed	16	0	0	0	7	1	0	24
	Normal static volumes	265	0	32	82	2	4959	36	5376
	Simple Restriction	57	21	0	0	0	120	140	338
$\Sigma$	1115	92	185	269	21	6188	270	8140	

Figura 5.16 - Tabela de contingência do XGboost na classificação da interpretação da PCT.

Tabela 5-12 - Resultados do desempenho dos modelos utilizados na classificação da interpretação da PCT.

Modelos ML	AUC	Precisão	F <sub>1</sub> score	Sensibilidade ( <i>precision</i> )	Valor preditivo positivo ( <i>recall</i> )
<b>XGradient Boosting</b>	0,884	0,751	0,731	0,729	0,751
<b>Random Forest</b>	0,847	0,735	0,718	0,712	0,735
<b>kNN</b>	0,809	0,721	0,703	0,697	0,721
<b>Logistic Regression</b>	0,864	0,701	0,698	0,715	0,701
<b>Neural Network</b>	0,785	0,704	0,623	0,573	0,704
<b>SVM</b>	0,685	0,550	0,577	0,644	0,550
<b>Naive Bayes</b>	0,791	0,495	0,537	0,658	0,495

O empilhamento dos modelos XGboost, kNN, LogR e NB gerou os seguintes resultados:

Tabela 5-13 - Resultados do desempenho do empilhamento de modelos na classificação da interpretação da PCT.

Modelos ML	AUC	Precisão	F <sub>1</sub> score	Sensibilidade ( <i>precision</i> )	Valor preditivo positivo ( <i>recall</i> )
<b>Stack</b>	0,881	0,709	0,715	0,734	0,709

## 6. Discussão

A análise exploratória dos dados permitiu identificar várias características dos indivíduos e as variáveis das suas espirometrias com potencial para serem introduzidas em modelos de ML, de forma a prever RV/TLC, RV e TLC.

Sabe-se que o sexo masculino e a altura de um indivíduo irão influenciar positivamente os seus volumes pulmonares, como foi observado nesta amostra. Em termos etários, a partir dos 35 anos verificou-se uma diminuição gradual da função respiratória no seu global, tal como esperado de acordo com a literatura existente.<sup>5</sup>

O peso foi outra característica antropométrica que, não se encontrando como variável nas equações de referência para os volumes pulmonares estáticos, apresentou uma elevada carga fisiopatológica.

Observou-se que com o aumento das categorias de IMC, o RV/TLC aumentou e o FEV<sub>1</sub> e a TLC diminuíram gradualmente, como era esperado dadas as implicações que a obesidade tem na função respiratória. No entanto, os indivíduos com baixo peso apresentaram valores mais elevados de RV/TLC e mais diminuídos de FEV<sub>1</sub> do que aqueles que apresentaram um peso ideal. Esta situação pode ser explicada como consequência do acréscimo do trabalho realizado pelos músculos respiratórios quando estão numa posição desfavorável em termos de eficiência energética. Isto acontece em doentes com DPOC com enfisema pulmonar marcado, em que a retificação da posição diafragmática conduz a um elevado consumo de energia na respiração a volume corrente, traduzindo-se numa diminuição do peso e do IMC do indivíduo.<sup>2,10,14,35</sup>

A assimetria de distribuição do FEV<sub>1</sub>/FVC e do RV/TLC pode ser justificada por se tratar de uma amostra com doentes pulmonares, nomeadamente indivíduos com DPOC com fenótipo de enfisema, e é possível que este desvio em relação à normalidade seja devido a *outliers* associados à evolução da gravidade da patologia, com obstrução das vias aéreas mais acentuada (FEV<sub>1</sub>/FVC muito reduzida) e com hiperinsuflação (RV/TLC muito aumentado).<sup>36</sup> Atendendo que a maioria dos exames avaliados neste estudo não apresentaram alterações significativas dos parâmetros espirométricos e pletismográficos, a presença de doentes mais graves leva à presença de *outliers*, provocando uma assimetria expectável da amostra.

Da análise entre as variáveis espirométricas e os volumes não mobilizáveis medidos por PCT, o FEV<sub>1</sub> apresentou uma correlação negativa forte com o RV/TLC. Isto

relaciona-se com o agravamento da obstrução das vias aéreas, condicionando uma limitação mais acentuada do fluxo expiratório, impedindo uma expiração completa e causando um aumento do RV.<sup>36</sup>

A correlação positiva forte entre a FVC e a TLC indica que quanto maior for a capacidade do indivíduo em mobilizar grandes quantidades de ar, maior será a dimensão estrutural do pulmão. Contudo, é necessário referir que tal não se aplica aos casos de hiperinsuflação grave, em que a limitação ao fluxo expiratório é tão acentuada que a FVC está diminuída, enquanto a TLC está bastante elevada.<sup>35</sup> Ainda assim, dada a prevalência de espirometrias sem alterações e de volumes estáticos normais, esta correlação positiva entre a FVC e a TLC não é inesperada.

A tentativa de prever ou classificar o RV/TLC, o RV e a TLC através dos modelos de ML gerou resultados globalmente satisfatórios.

Na comparação dos diferentes modelos na classificação dicotómica do RV/TLC (normal vs. aumentado), dos sete modelos testados, apenas o SVM apresentou uma AUC inferior a 0,8. O XGboost foi o modelo que melhor desempenho mostrou com uma AUC de 0,874, seguido da RF (0,841) e do kNN (0,808).

Dado que alguns modelos (LogR) apresentaram valores inferiores de erro tipo II (falsos negativos) e outros modelos (XGboost) apresentaram valores inferiores de erro tipo I (falsos positivos), procedeu-se ao empilhamento de vários modelos de diferentes características cujo resultado obteve uma AUC de 0,861. No entanto não se verificou uma melhoria significativa em relação ao modelo isolado com melhor desempenho, apesar de ter necessitado de mais tempo de treino para obter resultados.

Na tentativa de prever o valor do *z score* do RV/TLC, os modelos testados obtiveram resultados fracos, em que o melhor modelo foi a LR com o  $R^2$  de 0,442 e o erro médio 0,595. Dado que uma diferença do *z score* de 0,5 é clinicamente relevante, estes modelos foram insuficientes para fazer face ao problema proposto.<sup>5</sup>

Como os *z score* são uma forma de aferir o quão distante o parâmetro se encontra do valor previsto para um indivíduo saudável da mesma idade, sexo e altura, os modelos perdem o acesso a estas características e pequenas diferenças de *z score* podem representar diferenças significativas no valor absoluto do RV/TLC.

Neste sentido, fez-se a tentativa de prever o valor absoluto do RV/TLC, com valores de desempenho substancialmente melhores do que na predição dos *z score*. O

melhor algoritmo foi a LR, com um  $R^2$  de 0,752 (aceitável) e com RMSE de 0,051. O kNN teve um desempenho idêntico, ao contrário da NN que teve um desempenho fraco em comparação com os outros modelos.

Na previsão do RV, o melhor modelo foi LR, com um  $R^2$  de 0,539 (fraco), um RMSE de 0,479 e um MAE de 0,372 L, tendo sido o parâmetro com pior desempenho na medição dos valores absolutos (o RV médio foi de 2,17 L).

A previsão do valor absoluto de TLC foi aquela que apresentou melhores resultados nos problemas de regressão testados, com a LR a obter um  $R^2$  de 0,856 com um RMSE de 0,491. O MAE foi de 0,380 L para um parâmetro cuja média é de 5,93 L. A NN e o kNN apresentaram resultados idênticos à LR.

Por último, na classificação categórica dos volumes estáticos tendo em conta os parâmetros espirométricos, biológicos e antropométricos, o XGboost obteve uma boa AUC de 0,884 e um  $F_1$  score de 0,731, com os restantes modelos a produzirem desempenhos igualmente bons ou aceitáveis. De referir, no entanto, que grande parte dos indivíduos tinha volumes estáticos considerados normais, com uma prevalência de 66,0% do total da amostra. Quer isto dizer que se um algoritmo identificar toda a amostra como tendo volumes estáticos normais, a sua precisão será de 0,660. Ainda assim, o XGboost teve uma precisão satisfatória e superior, de 0,751.

Com o empilhamento da XGboost, kNN, LogR e NB, a AUC foi de 0,881 e o  $F_1$  score de 0,723, não traduzindo um melhor desempenho, apesar do aumento da complexidade do modelo, tal como já se tinha verificado no problema de classificação do RV/TLC.

Foi igualmente interessante verificar que algumas variáveis apresentaram um ganho de informação em relação às restantes. Em termos dos modelos de classificação, os parâmetros espirométricos basais (nomeadamente *z score* e percentagens do valor previsto) foram aqueles que tiveram um peso maior, enquanto os valores absolutos da espirometria após prova de BD foram os que ofereceram mais informação aos modelos de regressão.

Uma justificação para esta observação nos modelos classificativos pode dever-se à categorização da alteração do RV/TLC e das alterações pletismográficas ter por base o *z score* basal dos volumes pulmonares.

Por outro lado, a prova de resposta ao BD pode dar muita informação ao algoritmo de regressão devido à possível redistribuição de volumes estáticos, como consequência direta do aumento dos volumes mobilizáveis.

Na era da Medicina de precisão, antevê-se que a aliança entre ML e PFR assumirá um papel não só na melhoria da interpretação e sensibilidade das PFR, mas também na descoberta de padrões e biomarcadores que possibilitarão o tratamento precoce das doenças respiratórias.<sup>3-5</sup>

Topalovic e colaboradores<sup>37</sup> estudaram o desempenho de um modelo que treinou em 1500 casos de PFR com diagnóstico estabelecido. Compararam a classificação do modelo em 50 novas PFR com a interpretação de 120 pneumologistas. Os pneumologistas identificaram corretamente a alteração ventilatória em 74% das PFR e identificaram corretamente o diagnóstico em 45% dos casos. O modelo identificou corretamente a alteração ventilatória em 100% das PFR e o diagnóstico em 82% de todos casos. Com a continuação deste projeto<sup>38</sup>, observaram que unindo a interpretação conjunta dos pneumologistas com o modelo de ML, os resultados superiorizaram-se comparativamente a cada abordagem isoladamente.

Ioachimescu e colaboradores<sup>39-41</sup> aliaram um parâmetro pouco utilizado na interpretação da espirometria (a área abaixo da curva débito-volume) com algoritmos de ML e observaram que esta abordagem diferenciava corretamente as alterações ventilatórias.

Vários estudos foram publicados com a aplicação de modelos na detecção e progressão da DPOC<sup>42-47</sup>, da asma brônquica<sup>42,48-50</sup>, da esclerodermia<sup>51</sup>, de tumores pulmonares<sup>52-57</sup>, da hipertensão pulmonar<sup>58,59</sup>, das patologias do interstício<sup>60,61</sup> e do SARS-CoV<sub>2</sub>.<sup>62,63</sup>

Outros estudos previram resultados das PFR através de meios completos de diagnóstico imagiológicos<sup>64-66</sup>, da análise de voz<sup>67</sup> e até através de aplicações de *smartphone*.<sup>68</sup>

Alter e colaboradores<sup>69</sup> apresentaram um objetivo idêntico aos desta dissertação, em que tentaram prever o aumento do RV/TLC com base em 7157 registos de FEV<sub>1</sub>, FVC e FEV<sub>1</sub>/FVC. Neste estudo obtiveram igualmente bons resultados (AUC mais elevada de 0,889), contudo este foi realizado com uma seleção de indivíduos com uma patologia específica (DPOC).

A atual dissertação é, do conhecimento do seu autor, a única até à data que testou a capacidade de diferentes modelos de ML preverem diversos parâmetros (valores absolutos, *z score* e classificação categórica) de volumes pulmonares não mobilizáveis, com recurso a dados recolhidos de espirometrias, numa amostra recolhida em um hospital que acolhe um largo espectro de patologias com repercussão a nível respiratório.

## 7. Conclusão

A aplicação nos cuidados de saúde de técnicas de suporte à decisão clínica, como as estudadas nesta dissertação, permitirá não apenas aos profissionais de saúde uma melhor fundamentação das suas decisões, mas também tornar mais eficiente a acessibilidade dos doentes respiratórios a certos meios complementares de diagnóstico, evitando atrasos no diagnóstico ou tempos de espera elevados.

Esta dissertação apresenta algumas limitações, nomeadamente a inexistência de dados sobre a raça dos participantes, que conduz ao viés de considerar toda a amostra como caucasiana. Ainda assim, este condicionamento terá impacto apenas nos *z score* e percentagens do valor previsto nos parâmetros das PFR (com a sua possível subvalorização), não influenciando os valores absolutos ou as relações entre parâmetros (como é o caso de RV/TLC).

Outra limitação é a ausência de informação sobre os antecedentes pessoais, a histórica clínica ou os hábitos tabágicos dos participantes, parâmetros com relevância para a avaliação da função respiratória e possivelmente determinantes em termos de ganho de informação para o treino dos algoritmos de ML.

De igual forma, a falta de dados sobre os débitos expiratórios medidos na espirometria e sobre outros volumes tais como a IC e o VRE são limitações consideráveis que, se introduzidos no treino, poderão melhorar o desempenho dos modelos. Outro fator ao qual não houve acesso foi ao grau de qualidade de cada espirometria, pois exames com pouca qualidade apresentam pouca informação válida e induzirão em erro a aprendizagem automática computacional.

Neste projeto não foi realizada uma otimização exaustiva dos hiperparâmetros de cada modelo, pelo que é possível considerar que o desempenho destes poderá ser melhorado no futuro. Existe inclusivamente a possibilidade de que dada a variedade de algoritmos de ML que existem atualmente, modelos diferentes aos utilizados neste estudo poderão ser aplicados, com possível melhoria dos resultados. Por exemplo, dada a relação polinomial entre o FEV<sub>1</sub> e o RV/TLC, um modelo não linear poderá obter melhores resultados comparativamente à LR.

O desempenho no mundo real de qualquer modelo de ML está dependente de uma validação externa, pelo que como trabalho futuro poderá ser considerada a

aplicação destes modelos em bases de dados externas à população estudada, com uma otimização prévia dos hiperparâmetros dos algoritmos com melhor desempenho.

Este estudo focou-se somente em indivíduos adultos e idosos, no entanto não existe qualquer impedimento para a utilização no futuro desta mesma metodologia de forma a desenvolver modelos de ML para serem aplicados em crianças e adolescentes.

Outra área promissora no campo da função respiratória e da sua associação com ML é o treino de modelos de acordo com as imagens geradas nas PFR, como os gráficos da espirometria (débito-volume e volume-tempo) ou da PCT (resistência-volume). Esta metodologia poderá ter bons resultados não só na classificação da alteração ventilatória, mas também na identificação de padrões associados a patologias específicas, pois a quantidade de informação contida nestas imagens pode superiorizar-se aos valores discretos que constam em bases de dados tabulares.

Assim, é possível perspetivar um futuro em que um paciente com sintomas que indiquem a possibilidade de alguma disfunção respiratória, no âmbito de uma consulta de Medicina Geral e Familiar, Pneumologia ou Imunoalergologia, faça uma espirometria e, com esta, seja possível extrapolar a probabilidade da existência de alteração dos volumes pulmonares estáticos. Este futuro só será possível se no presente se continuarem a realizar PFR de elevada qualidade, para que os dados contidos nelas sejam o mais confiáveis possível.

Concluindo, esta dissertação mostra que é possível prever com boa precisão os volumes pulmonares não mobilizáveis medidos em PCT, tendo em consideração dados biológicos, antropométricos e espirométricos e utilizando técnicas da ciência de dados para a produção de modelos de ML fidedignos.

## 8. Bibliografia

1. Soriano JB, Kendrick PJ, Paulson KR, Gupta V, Abrams EM, Adedoyin RA, et al. Prevalence and attributable health burden of chronic respiratory diseases, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet Respir Med*. 2020 Jun 1;8(6):585–96.
2. Mottram C. *Ruppel's Manual of Pulmonary Function Testing*. 10th ed. Missouri: Elsevier; 2013.
3. Khemasuwan D, Sorensen JS, Colt HG. Artificial intelligence in pulmonary medicine: Computer vision, predictive model and covid-19. *European Respiratory Review*. 2020 Sep 30;29(157):1–16.
4. Giri PC, Chowdhury AM, Bedoya A, Chen H, Lee HS, Lee P, et al. Application of Machine Learning in Pulmonary Function Assessment Where Are We Now and Where Are We Going? *Front Physiol*. 2021 Jun 24;12.
5. Stanojevic S, Kaminsky DA, Miller MR, Thompson B, Aliverti A, Barjaktarevic I, et al. ERS/ATS technical standard on interpretive strategies for routine lung function tests. *European Respiratory Journal*. 2022 Jul 1;60(1).
6. Kaminsky DA, Irvin CG. *Pulmonary Function Testing*. Kaminsky DA, Irvin CG, editors. Cham: Springer International Publishing; 2018. (Respiratory Medicine).
7. Tseng HJ, Henry TS, Veeraraghavan S, Mittal PK, Little BP. Pulmonary function tests for the radiologist. *Radiographics*. 2017 Jul 1;37(4):1037–58.
8. Graham BL, Steenbruggen I, Barjaktarevic IZ, Cooper BG, Hall GL, Hallstrand TS, et al. Standardization of spirometry 2019 update an official American Thoracic Society and European Respiratory Society technical statement. Vol. 200, *American Journal of Respiratory and Critical Care Medicine*. American Thoracic Society; 2019. p. E70–88.
9. West JB. *Respiratory Physiology The Essentials*. 9th ed. Baltimore: Lippincott Williams & Wilkins; 2012.
10. Leith D, Brown R. Human lung volumes and the mechanisms that set them. *Eur Respir J*. 1999;13:468–72.

11. Criée CP, Sorichter S, Smith HJ, Kardos P, Merget R, Heise D, et al. Body plethysmography - Its principles and clinical use. Vol. 105, Respiratory Medicine. 2011. p. 959–71.
12. Hall GL, Filipow N, Ruppel G, Okitika T, Thompson B, Kirkby J, et al. Official ERS technical standard: Global lung function initiative reference values for static lung volumes in individuals of european ancestry. European Respiratory Journal. 2021;57(3).
13. O'Donnell DE, Laveneziana P. Physiology and consequences of lung hyperinflation in COPD. European Respiratory Review. 2006 Dec 1;15(100):61–7.
14. Salome CM, King GG, Berend N. Physiology of obesity and effects on lung function. J Appl Physiol. 2010 Jan;108(1):206–11.
15. Graham BL, Brusasco V, Burgos F, Cooper BG, Jensen R, Kendrick A, et al. 2017 ERS/ATS standards for single-breath carbon monoxide uptake in the lung. European Respiratory Journal. 2017 Jan 1;49(1).
16. Stanojevic S, Graham BL, Cooper BG, Thompson BR, Carter KW, Francis RW, et al. Official ERS technical standards: Global Lung Function Initiative reference values for the carbon monoxide transfer factor for Caucasians. European Respiratory Journal. 2017 Sep 11;50(3):1700010.
17. Quanjer PH, Stanojevic S, Cole TJ, Baur X, Hall GL, Culver BH, et al. Multi-ethnic reference values for spirometry for the 3-95-yr age range: The global lung function 2012 equations. European Respiratory Journal. 2012 Dec 1;40(6):1324–43.
18. Medical International Research - MIR. Interpreting results with GLI-2012 using Z-score and LLN. <https://spirometry.com/en/news/interpreting-results-with-gli-2012-using-z-score-and-lln/>. 2020.
19. Uddin S, Khan A, Hossain ME, Moni MA. Comparing different supervised machine learning algorithms for disease prediction. BMC Med Inform Decis Mak. 2019 Dec 21;19(1).
20. Io Vercio L, Amador K, Bannister JJ, Crites S, Gutierrez A, MacDonald ME, et al. Supervised machine learning tools: a tutorial for clinicians. J Neural Eng. 2020 Dec 22;17(6):062001.
21. Tohka J, van Gils M. Evaluation of machine learning algorithms for health and wellness applications: A tutorial. Comput Biol Med. 2021 May 1;132.

22. Moorthie S, Babb de Villiers C, Brigden T, Gaynor L, Hall A, Johnson E, et al. Polygenic scores, risk and cardiovascular disease. 2019.
23. James G, Witten D, Hastie T, Tibshirani R. An Introduction to Statistical Learning with Applications in R. Casella G, Fienberg S, Olkin I, editors. New York: Springer; 2017.
24. Sontakke SA, Lohokare J, Dani R, Shivagaje P. Classification of Cardiotocography Signals Using Machine Learning. In 2019. p. 439–50.
25. Bruce P, Bruce A, Gedeck P. Practical Statistics for Data Scientists. 2nd ed. Tache N, editor. Sebastopol: O'Reilly Media, Inc.; 2020.
26. Gupta T. Understanding Linear Regression [Internet]. 2021 [cited 2022 Sep 25]. Available from: <https://medium.com/codex/understanding-linear-regression-402df5a4667e>
27. Essampally D. Logistic Regression [Internet]. 2020 [cited 2022 Sep 25]. Available from: <https://dev86essampally.medium.com/logistic-regression-56b53b2fd7ab>
28. Valigi N. Naive Bayes classifiers in TensorFlow [Internet]. [cited 2022 Sep 25]. Available from: <https://nicolovaligi.com/articles/naive-bayes-tensorflow/>
29. Chouinard JC. k-Nearest Neighbors (KNN) in Python [Internet]. 2022 [cited 2022 Sep 25]. Available from: <https://www.jcchouinard.com/k-nearest-neighbors/>
30. Zhu Z. Explain Support Vector Machines in Mathematic Details [Internet]. 2020 [cited 2022 Sep 25]. Available from: <https://towardsdatascience.com/explain-support-vector-machines-in-mathematic-details-c7cc1be9f3b9>
31. Martin Pa. The Universal Approximation Theorem is Terrifying [Internet]. 2022 [cited 2022 Sep 25]. Available from: <https://medium.com/@patrickmartinaz/the-universal-approximation-theorem-is-terrifying-83a53acc4192>
32. Theja Y, Liming B, Xuanchang L, Longhao Z, Balaji A. Which is the real money maker, Random Forest regressors or LSTM Networks? [Internet]. 2020 [cited 2022 Sep 25]. Available from: <https://medium.com/@nusfintech.ml/which-is-the-real-money-maker-random-forest-regressors-or-lstm-networks-a2153c0f8e92>
33. Deng H, Zhou Y, Wang L, Zhang C. Ensemble learning for the early prediction of neonatal jaundice with genetic features. BMC Med Inform Decis Mak. 2021 Dec 1;21(1):338.

34. Global Lung Function Initiative. Global Lung Function Initiative calculators for Spirometry, TLCO and Lung volume [Internet]. 2021 [cited 2022 Sep 25]. Available from: <http://gli-calculator.ersnet.org/index.html>
35. Ferguson GT. Why does the lung hyperinflate? In: Proceedings of the American Thoracic Society. 2006. p. 176–9.
36. West JB. Pulmonary Pathophysiology The Essentials. 8th ed. Baltimore: Lippincott Williams & Wilkins;
37. Topalovic M, Das N, Burgel PR, Daenen M, Derom E, Haenebalcke C, et al. Artificial intelligence outperforms pulmonologists in the interpretation of pulmonary function tests. *European Respiratory Journal*. 2019 Apr 1;53(4).
38. Das N, Happaerts S, Topalovic M, Janssens W. Pulmonologists collaborate with explainable artificial intelligence for superior interpretation of pulmonary function tests. In: m-Health/e-health. European Respiratory Society; 2021. p. PA3443.
39. Ioachimescu OC, Stoller JK. An alternative spirometric measurement : Area under the expiratory flow–volume curve. *Ann Am Thorac Soc*. 2020 May 1;17(5):582–8.
40. Ioachimescu OC, Stoller JK. Area Under the Expiratory Flow–Volume Curve (AEX): Assessing Bronchodilator Responsiveness. *Lung*. 2020 Jun 1;198(3):471–80.
41. Ioachimescu OC, Stoller JK, Garcia-Rio F. Area under the expiratory flow-volume curve: predicted values by artificial neural networks. *Sci Rep*. 2020 Dec 1;10(1).
42. Spathis D, Vlamos P. Diagnosing asthma and chronic obstructive pulmonary disease with machine learning. *Health Informatics J*. 2019 Sep 1;25(3):811–27.
43. Amaral JLM, Lopes AJ, Jansen JM, Faria ACD, Melo PL. Machine learning algorithms and forced oscillation measurements applied to the automatic identification of chronic obstructive pulmonary disease. *Comput Methods Programs Biomed*. 2012 Mar;105(3):183–93.
44. Chen W, Sin DD, FitzGerald JM, Safari A, Adibi A, Sadatsafavi M. An Individualized Prediction Model for Long-term Lung Function Trajectory and Risk of COPD in the General Population. In: *Chest*. Elsevier Inc; 2020. p. 547–57.
45. Kanwade A, Bairagi VK. Classification of COPD and normal lung airways using feature extraction of electromyographic signals. *Journal of King Saud University - Computer and Information Sciences*. 2019 Oct 1;31(4):506–13.

46. Bodduluri S, Nakhmani A, Reinhardt JM, Wilson CG, McDonald ML, Rudraraju R, et al. Deep neural network analyses of spirometry for structural phenotyping of chronic obstructive pulmonary disease. *JCI Insight*. 2020 Jul 9;5(13).
47. Bodduluri S, Nakhmani A, Reinhardt JM, Wilson CG, McDonald ML, Rudraraju R, et al. Deep neural network analyses of spirometry for structural phenotyping of chronic obstructive pulmonary disease. *JCI Insight*. 2020 Jul 9;5(13).
48. Tomita K, Nagao R, Touge H, Ikeuchi T, Sano H, Yamasaki A, et al. Deep learning facilitates the diagnosis of adult asthma. *Allergology International*. 2019 Oct 1;68(4):456–61.
49. Feng Y, Wang Y, Zeng C, Mao H. Artificial intelligence and machine learning in chronic airway diseases: Focus on asthma and chronic obstructive pulmonary disease. Vol. 18, *International Journal of Medical Sciences*. Ivyspring International Publisher; 2021. p. 2871–89.
50. Lovrić M, Banić I, Lacić E, Pavlović K, Kern R, Turkalj M. Predicting treatment outcomes using explainable machine learning in children with asthma. *Children*. 2021 May 1;8(5).
51. Murdaca G, Caprioli S, Tonacci A, Billeci L, Greco M, Negrini S, et al. A machine learning application to predict early lung involvement in scleroderma: A feasibility evaluation. *Diagnostics*. 2021 Oct 1;11(10).
52. Ardila D, Kiraly AP, Bharadwaj S, Choi B, Reicher JJ, Peng L, et al. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nat Med*. 2019 Jun 20;25(6):954–61.
53. Baldwin DR, Gustafson J, Pickup L, Arteta C, Novotny P, Declerck J, et al. External validation of a convolutional neural network artificial intelligence tool to predict malignancy in pulmonary nodules. *Thorax*. 2020 Apr;75(4):306–12.
54. Massion PP, Antic S, Ather S, Arteta C, Brabec J, Chen H, et al. Assessing the Accuracy of a Deep Learning Method to Risk Stratify Indeterminate Pulmonary Nodules. *Am J Respir Crit Care Med*. 2020 Jul 15;202(2):241–9.
55. Ciompi F, Chung K, van Riel SJ, Setio AAA, Gerke PK, Jacobs C, et al. Towards automatic pulmonary nodule management in lung cancer screening with deep learning. *Sci Rep*. 2017 Jun 16;7(1):46479.

56. Nam JG, Park S, Hwang EJ, Lee JH, Jin KN, Lim KY, et al. Development and Validation of Deep Learning–based Automatic Detection Algorithm for Malignant Pulmonary Nodules on Chest Radiographs. *Radiology*. 2019 Jan;290(1):218–28.
57. Zhao W, Yang J, Sun Y, Li C, Wu W, Jin L, et al. 3D Deep Learning from CT Scans Predicts Tumor Invasiveness of Subcentimeter Pulmonary Adenocarcinomas. *Cancer Res*. 2018 Dec 15;78(24):6881–9.
58. Sweatt AJ, Hedlin HK, Balasubramanian V, Hsi A, Blum LK, Robinson WH, et al. Discovery of Distinct Immune Phenotypes Using Machine Learning in Pulmonary Arterial Hypertension. *Circ Res*. 2019 Mar 15;124(6):904–19.
59. Leha A, Hellenkamp K, Unsöld B, Mushemi-Blake S, Shah AM, Hasenfuß G, et al. A machine learning approach for the prediction of pulmonary hypertension. *PLoS One*. 2019 Oct 25;14(10):e0224453.
60. Walsh SLF, Calandriello L, Silva M, Sverzellati N. Deep learning for classifying fibrotic lung disease on high-resolution computed tomography: a case-cohort study. *Lancet Respir Med*. 2018 Nov;6(11):837–45.
61. Christe A, Peters AA, Drakopoulos D, Heverhagen JT, Geiser T, Stathopoulou T, et al. Computer-Aided Diagnosis of Pulmonary Fibrosis Using Deep Learning and CT Images. *Invest Radiol*. 2019 Oct;54(10):627–32.
62. Wang S, Kang B, Ma J, Zeng X, Xiao M, Guo J, et al. A deep learning algorithm using CT images to screen for Corona virus disease (COVID-19). *Eur Radiol*. 2021 Aug 24;31(8):6096–104.
63. Li L, Qin L, Xu Z, Yin Y, Wang X, Kong B, et al. Using Artificial Intelligence to Detect COVID-19 and Community-acquired Pneumonia Based on Pulmonary CT: Evaluation of the Diagnostic Accuracy. *Radiology*. 2020 Aug;296(2):E65–71.
64. Gawlitza J, Sturm T, Spohrer K, Henzler T, Akin I, Schönberg S, et al. Predicting pulmonary function testing from quantified computed tomography using machine learning algorithms in patients with COPD. *Diagnostics*. 2019 Mar 1;9(1).
65. Schroeder JD, Lanfredi RB, Li T, Chan J, Vachet C, Paine R, et al. Prediction of Obstructive Lung Disease from Chest Radiographs via Deep Learning Trained on Pulmonary Function Data. *International Journal of COPD*. 2020;15:3455–66.
66. Ramalho GLB, Filho PPR, de Medeiros NSF, Cortez PC. Lung disease detection using feature extraction and extreme learning machine. *Revista Brasileira de Engenharia Biomedica*. 2014;30(3):207–14.

67. Alam MZ, Simonetti A, Brillantino R, Tayler N, Grainge C, Siribaddana P, et al. Predicting Pulmonary Function From the Analysis of Voice: A Machine Learning Approach. *Front Digit Health*. 2022 Feb 8;4.
68. Teixeira J. Signal processing and machine learning methods and algorithm analysis for classification of lung function on a smartphone app. [Porto]: Faculdade de Engenharia da Universidade do Porto; 2014.
69. Alter P, Orszag J, Kellerer C, Kahnert K, Speicher T, Watz H, et al. Prediction of air trapping or pulmonary hyperinflation by forced spirometry in COPD patients: results from COSYCONET. *ERJ Open Res*. 2020 Jul;6(3):00092–2020.

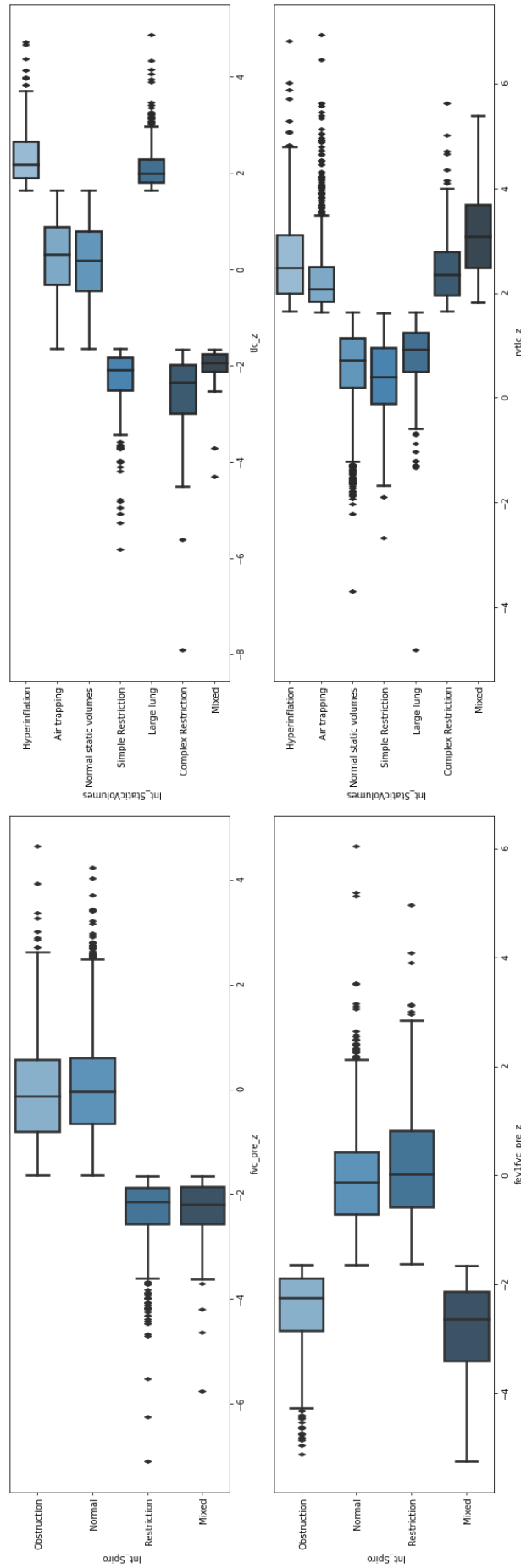


Figura 0.1 – Box plots da distribuição dos z score de FVC e FEV<sub>1</sub>/FVC na interpretação da espirometria e da TLC e RV/TLC para a PCT.

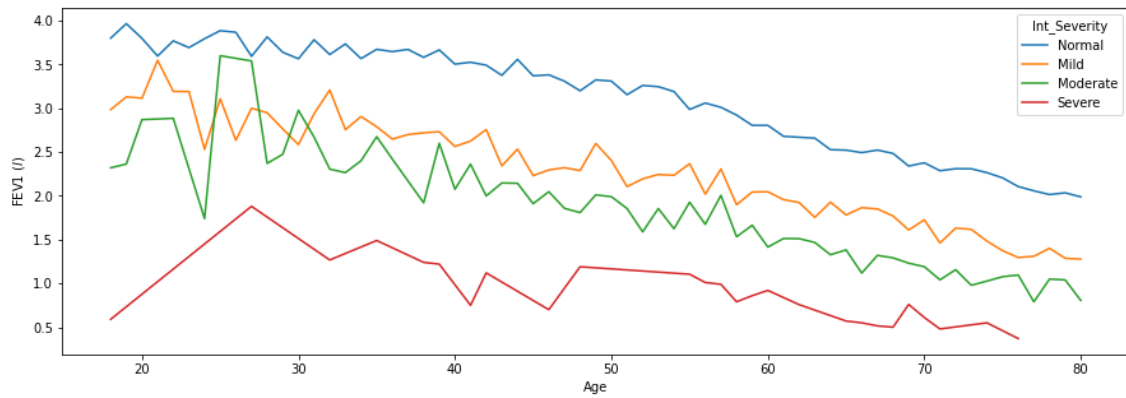


Figura 0.2 - Variação do z score do FEV<sub>1</sub> ao longo da idade dos participantes e de acordo com a gravidade da alteração ventilatória.

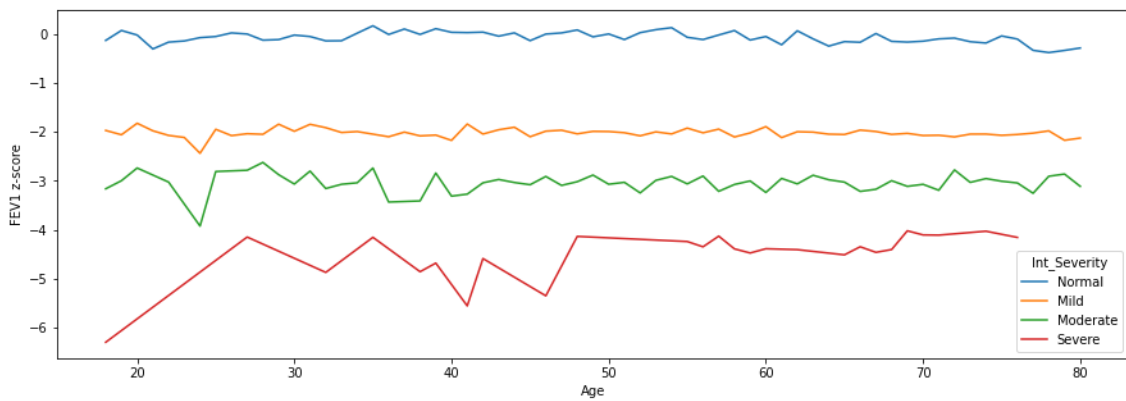


Figura 0.3 - Variação do valor absoluto de RV/LC ao longo da idade dos participantes e de acordo com a gravidade da alteração ventilatória.

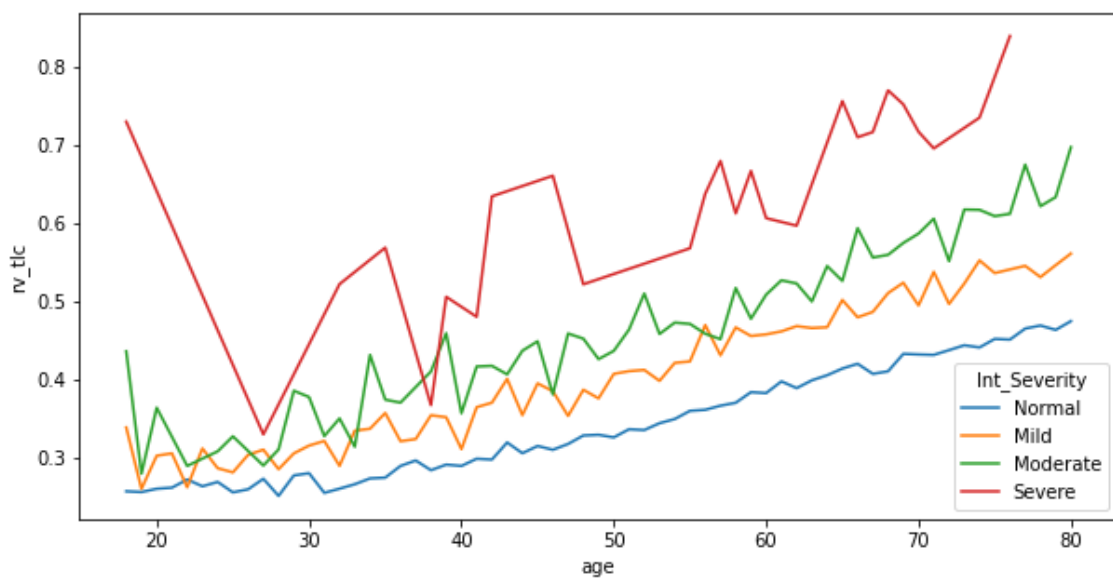


Figura 0.4 - Variação do valor absoluto do FEV<sub>1</sub> ao longo da idade dos participantes e de acordo com a gravidade da alteração ventilatória.

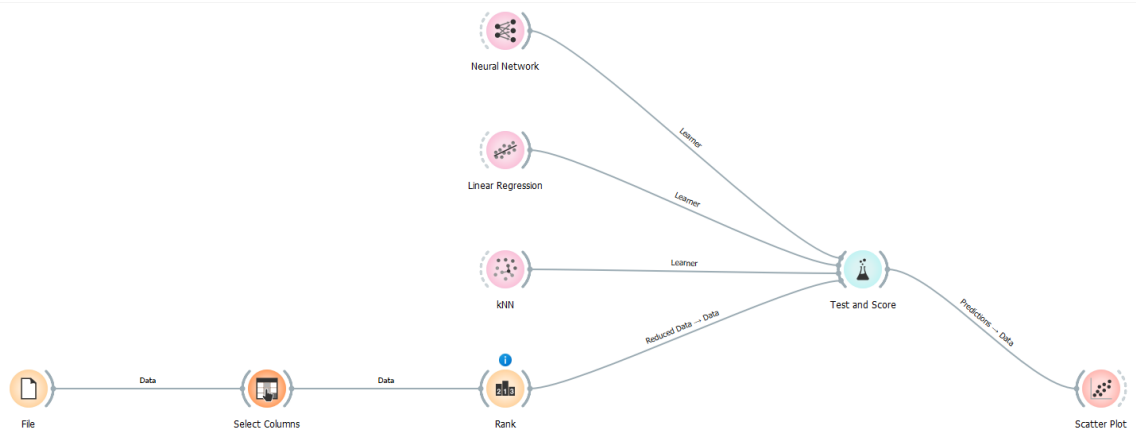


Figura 0.5 - Esquema utilizado no software Orange para o treino e avaliação do desempenho dos modelos de regressão.

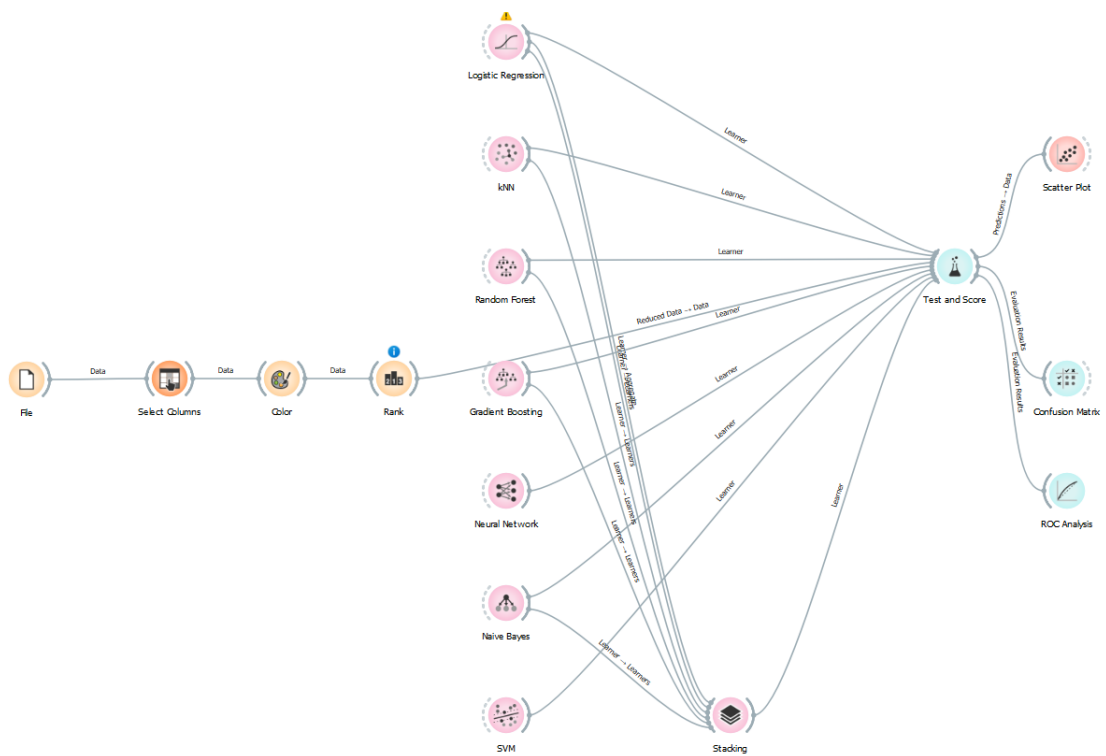


Figura 0.6 - Esquema utilizado no software Orange para o treino e avaliação do desempenho dos modelos de classificação.