



M 2014

CLASSIFICATION OF LUNG FUNCTION ON A SMARTPHONE APP

JOÃO PEDRO FONSECA TEIXEIRA

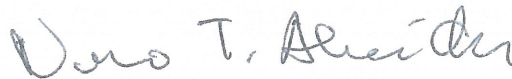
DISSERTAÇÃO DE Mestrado APRESENTADA
À FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO EM
ENGENHARIA ELECTROTÉCNICA E DE COMPUTADORES

A Dissertação intitulada

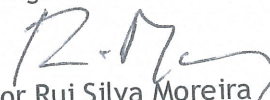
“Classification of Lung Function on a Smartphone App”

foi aprovada em provas realizadas em 25-07-2014


o júri



Presidente Professor Doutor José Nuno Teixeira de Almeida
Professor Auxiliar do Departamento de Engenharia Eletrotécnica e de Computadores
da Faculdade de Engenharia da Universidade do Porto

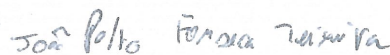


Professor Doutor Rui Silva Moreira
Professor Associado do Faculdade de Ciências e Tecnologia da Universidade
Fernando Pessoa



Professor Doutor Luís Filipe Pinto de Almeida Teixeira
Professor Auxiliar do Departamento de Engenharia Informática da Faculdade de
Engenharia da Universidade do Porto

O autor declara que a presente dissertação (ou relatório de projeto) é da sua exclusiva autoria e foi escrita sem qualquer apoio externo não explicitamente autorizado. Os resultados, ideias, parágrafos, ou outros extratos tomados de ou inspirados em trabalhos de outros autores, e demais referências bibliográficas usadas, são corretamente citados.



Autor - João Pedro Fonseca Teixeira

FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO



Classification of Lung Function on a Smartphone App

João Pedro Fonseca Teixeira

Mestrado Integrado em Engenharia Eletrotécnica e de Computadores

FEUP Supervisor: Luís Teixeira

External Supervisor: Miguel Coimbra

July 31, 2014

Resumo

Asma e DPOC (Doença Pulmonar Obstrutiva Crónica) são doenças crónicas do foro respiratório que afetam dezenas de milhares de pessoas por todo o mundo. Podem provocar falta de ar, diminuição severa da qualidade de vida e, no caso da DPOC, pode ser causa de morte, se não for detetada e devidamente controlada.

As abordagens e aparelhos Mobile Health (mHealth) têm aumentado em popularidade e estão a revolucionar os sistemas de saúde mundiais, através de formas mais baratas e eficientes para prevenção, diagnóstico e monitorização de doenças e respetivos tratamentos.

Nesta dissertação, pretende-se chegar à melhor e à mais eficiente combinação dos métodos de processamento de sinal e aprendizagem computacional estudados, de modo a poder ser incluída numa aplicação de *smartphone* para classificar a função respiratória, baseando-se apenas em gravações de microfone. A possibilidade de produzir semelhante aplicação poderá levar ao aumento da prevenção e deteção precoce de dificuldades respiratórias e proporcionar um sistema de monitorização de doenças acessível, portátil e de reduzidas dimensões.

O projeto foi dividido em duas partes. A primeira consiste no processamento de sinal, em que o sinal de áudio é processado e são calculadas várias envolventes, de onde se extraem os parâmetros clínicos de espirometria. A segunda parte consiste nos modelos aprendizagem computacional que permitem regredir os parâmetros e classificar as gravações, seja no par de classes Normal e Anormal seja no conjunto de classes função respiratória Normal e alterações ventilatórias Obstrutiva, Restritiva e Mista.

O conjunto de dados é composto por 101 gravações de manobras de expiração forçada, de 61 pacientes. A comparação de experiências de processamento de sinal foi realizada numa abordagem em *backward selection*, reduzindo de um total inicial de 54 envolventes, para 8 envolventes finais, por gravação. Geraram-se modelos de regressão para *Peak Expiratory Flow* (PEF), *Forced Expiratory Volume* em menos de 1 segundo (FEV_1), *Forced Volume Capacity* (FVC) e índice FEV_1/FVC , obtendo os erros médios de cerca de 21.3%, 32.2%, 22.6% e 10.0%, respetivamente. Todos os modelos de aprendizagem computacional foram treinados e testados num esquema de validação cruzada em 5 *fold*, com um conjunto de dados que continha a idade e a altura dos pacientes. O modelo de classificação para duas classes apresentou um erro de classificação de 8%, com precisão e *recall* de 90.90% e 98.59%, respetivamente. O modelo de classificação em múltiplas classes apresentou um erro de classificação de cerca de 10%.

Os resultados experimentais deste projeto são promissores e encorajam futuros desenvolvimentos do sistema.

Abstract

Asthma and COPD (Chronic Obstructive Pulmonary Disease) are chronic lung conditions that affect tens of millions of people worldwide. They can result in breathlessness, a harsh decrease in quality of life and, in case of COPD, death, if not detected and duly managed.

Mobile Health (mHealth) approaches and devices have been increasing in popularity and are revolutionizing the world's health care systems by providing cheaper and more efficient ways to prevent, diagnose and monitor diseases, as well as to track a patient's treatment response.

In this dissertation, we aim to find the best and most efficient combination of signal processing and machine learning methods in order to produce a smartphone app that could accurately classify lung function, using microphone recordings as the only input. The possibility to produce such application is bound to increase prevention and early detection of respiratory diseases and provide a highly portable and affordable lung monitoring system.

The project was divided into two procedural stages. The first consists of the signal processing stage, in which the audio signal is fed to several functions and transformed into an envelope, from which spirometry clinical parameters are extracted. The second consists of machine learning models that regress the parameters and classify the recordings, either into Normal versus Abnormal labels or into Normal lung function, Obstructive, Restrictive or Mixed abnormality labels.

The dataset was composed by 101 recordings of forced expiration maneuvers from 61 patients. The signal processing comparison experiments were conducted in a backward selection approach, reducing to 8 final envelopes, per recording, out of the initial 54, and produced a final regression model for *Peak Expiratory Flow* (PEF), *Forced Expiratory Volume* under 1 second (FEV_1), *Forced Volume Capacity* (FVC) e the index FEV_1/FVC , with mean error of around 21.3%, 32.2%, 22.6% and 10.0%, respectively. All machine learning models were trained and tested on a 5 fold cross validation scheme, with a dataset's feature space that contained the patients' age and height. The two label classification model presented an 8% misclassification rate with precision and recall values of 90.90% and 98.59%, respectively. The multiple label classification model presented a misclassification rate around 10%.

The experimental results from this project seem promising and encourage further development of the system.

Acknowledgements

First and foremost, I would like to thank my parents for their relentless support, care, enabling me to get here and for letting me become who I am.

I would like to thank the Faculty of Engineering of Porto (FEUP) and specially the Department of Electrical and Computer Engineering (DEEC) for their great resources and professionals, some of which I hold dear for shaping my appreciation and enthusiasm for engineering and broadening my horizons.

Thanks to professor João Fonseca for the opportunity and trust with the project.

Thanks to my supervisor, professor Luís Teixeira for his dedication and precious discussions.

Thanks to my supervisor, professor Miguel Coimbra for his availability and for providing valuable insight.

Thanks to Tiago Jacinto, Msc, Respiratory Physiologist at the lung function laboratory of the CUF Porto Institute, and researcher at CINTESIS, for being one of the founders of this project, and for providing insight on spirometry and lung function tests.

Thanks to the people at CINTESIS for their warm welcome, and specially to Ivânia Gonçalves, Rita Silva, Daniela Santos for their hard work and professionalism with the recordings and to Ana Sousa and Rita Amaral for their friendliness and support.

Thanks to Joana Santos for her love, patience and unyielding support.

Thanks to Bernardo Pinho for his friendship and working on his free time, on short notice.

Thanks to Francisco Martins for his companionship, pro tips and assistance.

Thanks to all my friends for the laughs off.

João Teixeira

*“Some people see things that are and ask, Why?
Some people dream of things that never were and ask, Why not?
Some people have to go to work and don’t have time for all that.”*

George Carlin

Contents

1	Introduction	1
1.1	Motivation	1
1.2	Problem Statement	2
1.2.1	Project Objectives	2
1.2.2	Application Requirements	3
1.3	The Company	3
1.4	Structure of the document	4
2	Literature Review	5
2.1	Spirometry	5
2.1.1	The science of lung function	5
2.1.2	Lung function parameters	7
2.2	Spirometry Technology	8
2.2.1	Pneumotachometers	8
2.2.2	Peak Flow Meters	11
2.2.3	Plethysmographs	11
2.3	eHealth and mHealth approaches	12
2.3.1	Games and Applications	12
2.3.2	Prototypes and developing projects	14
2.4	Software Solution Review	15
2.4.1	SpiroSmart	15
2.4.2	SpiroApp	15
2.4.3	mCOPD	15
3	System for Lung Function Classification	17
3.1	Initial System Architecture	17
3.1.1	Concept and Structure	17
3.2	Signal Processing	19
3.2.1	Pre-processing	19
3.2.2	Mean of Resonances	23
3.2.3	Hilbert Transform	23
3.2.4	Shannon Envelopes	25
3.2.5	Linear Predictive Coding	26
3.2.6	Post processing	28
3.3	Feature Extraction	28
3.4	Machine Learning	30
3.4.1	Regression Algorithms	30
3.4.2	Classification Algorithms	30

3.4.3	Ensemble Algorithms	33
4	Experimental Evaluations	35
4.1	Experimental Setup	35
4.1.1	Technologies and tools	35
4.1.2	Recording process/protocol	35
4.2	Dataset	36
4.3	Regression Experiments	37
4.3.1	Preliminary Regression Results	38
4.3.2	FEV ₁ /FVC Regression and Calculation	38
4.3.3	Comparison of Conversion vocal models	39
4.3.4	Filtering options	41
4.3.5	Hilbert Transform and Shannon Envelopes	43
4.3.6	Shannon Entropy and Energy	44
4.4	Additional Regression Experiments	46
4.4.1	LPC fundamental vocal complexity	46
4.4.2	Low Pass Filtering and Savitzky-Golay Smoothing	48
4.4.3	Polynomial Fitting	50
4.4.4	Final Regression Model	51
4.5	Two Label Classification	53
4.5.1	Tree Ensemble Setup	53
4.5.2	Classification with the 9 clinical parameters	53
4.5.3	Classification with the 4 most used clinical parameters	56
4.5.4	Influence of Height and Age in Classification feature space	58
4.6	Multiple Label Classification	59
4.6.1	Classification with the 9 clinical parameters	59
4.6.2	Classification with the 4 most popular clinical parameters	60
4.6.3	Influence of Height and Age in Classification feature space	61
4.6.4	Classification with the 4 most popular clinical parameters plus Height and Age	62
4.7	Final Prototype	63
5	Conclusion and Discussion	65
5.1	Conclusion	65
5.2	Discussion	66
5.2.1	Main problems	66
5.2.2	Future Work	67
A	Initial Method Discussion	71
A.1	Choosing the Signal Processing methods	71
A.2	Choosing the Machine Learning methods	72
A.2.1	Regression	72
A.2.2	Classification	73
	References	75

List of Figures

2.1	Respiratory Tidal Volume and Forced exhalation	6
2.2	Types of ventilatory defect with typical spiograms and flow-volume curves. Image taken from [1]	7
2.3	Common Flow Sensing Devices (Pneumotachometers). A-Fleisch type sensor, B-Heated wire sensor, C-Pitot tube sensor, D-Turbine sensor, E-Ultrasonic sensor. Image taken from [2].	10
3.1	Initial Signal Processing stage Diagram	18
3.2	Zero time Back Extrapolation algorithm. Top left: Stage 1, original Flow-Time envelope, grossly separating the inspiration from the expiration portions. Top right: Stage 2, Flow-Time Envelope with the inspiration portion removed. Bottom left: Stage 3, Volume-Time envelope back extrapolating the zero time. Bottom right: Stage 4, Flow-Time envelope with pre-zero time values removed.	20
3.3	Stage 3 of the Zero time Back Extrapolation algorithm (enlargement). The zero time is where the PEF time tangential line intercepts with the zero volume. The PEF time is marked as a circle.	20
3.4	Automatic crop steps. Top image: Original recording. Middle image: Recording after crop from the back extrapolation algorithm. Bottom image: Recording after the previous crop and the final, moving window crop.	21
3.5	Pre-processing conversion models' results	22
3.6	Frequency tracking spectrogram process, from left to right, original, noise filtered, persistence of resonances	24
3.7	Frequency tracking spectrogram output, on the left using bin tolerance of 0, on the right using bin tolerance of 2	24
3.8	Original signal (in black) and Hilbert transform envelope (in red): Top image is an unprocessed result. Bottom image is the filtered result	25
3.9	Comparison of different envelope methods. The solid line is the Shannon Energy that converts the original amplitude values and enhances the upper amplitudes. The dotted line is the Shannon Entropy that converts the amplitude values and enhances the lower amplitudes. The dashdot line is the absolute value ($ x $) that has a linear amplitude response. The dashed line is the Energy (x^2) which is non-linear and boosts the dynamic range of the signal.	26
3.10	Comparison of Original signal and Shannon envelopes (black) and filtered results (red)	27
3.11	Clinical curves	29
3.12	Regression Tree example, regressing PEF	31
3.13	Separable classification with Radial Basis kernel functions in different spaces. Left: original space. Right: feature space ¹	32

4.1	Screenshots from the enhanced recording app	36
4.2	FEV ₁ /FVC index error comparison, to the left, directly regressed values, and to the right, calculated based on regressed values of FEV ₁ and FVC, independently	39
4.3	Comparison of Bagging Regression Error Distribution along the clinical parameters and sets of methods (enlargement)	40
4.4	Split count distribution over the 9 parameters, for an envelope set containing all methods	42
4.5	Comparison of Low Pass Filter and Moving Average, Bagging Regression results. Error Distribution, throughout the clinical parameters, using the same set of methods (enlargement)	43
4.6	Comparison of Shannon envelopes and Hilbert Transform envelopes' influences on Bagging Regression results. Error Distribution, throughout the clinical parameters (enlargement)	44
4.7	Comparison of Shannon Energy and Entropy envelopes' influences on Bagging Regression results. Error Distribution, throughout the clinical parameters (enlargement)	45
4.8	Split count distribution over the 9 parameters, for a set containing both of Shannon envelope methods	45
4.9	Comparison of subsets of LPC envelopes' and their influences on Bagging Regression results. Error Distribution, throughout the clinical parameters (enlargement)	47
4.10	Split count distribution over the 9 parameters, for LPC set B (containing only orders 2 and 32)	47
4.11	Comparison of the Low Pass filter and Savitzky-Golay envelopes' and their influences on Bagging Regression results. Error Distribution, throughout the clinical parameters (enlargement)	49
4.12	Split count distribution over the 9 parameters, for the Savitzky-Golay and Low Pass filtering options (using both)	49
4.13	Comparison of the Low Pass filter and Polynomial Fitting envelopes' and their influences on Bagging Regression results. Error Distribution, throughout the clinical parameters (enlargement)	51
4.14	Comparison between a filtered envelope and a polynomial fitted envelope	51
4.15	Signal Processing stage Diagram	52
4.16	Signal Processing stage Final Diagram	57

List of Tables

2.1	Clinical parameters used	8
2.2	Microphone based smartphone solutions' characteristics	16
4.1	Recordings' Distribution and ground truth device	37
4.2	Recordings' clinical classification distribution	37
4.3	Regressed Values of FEV ₁ , using Bagging and Random Forests, using the original signal processing architecture	38
4.4	Conversion Models experiment - Regressed Values Error - Bagging	39
4.5	Conversion Models experiment - Regressed Values Error - Random Forest	40
4.6	Filtering options experiment - Regressed Values Error - Bagging	41
4.7	Filtering options experiment - Regressed Values Error - Random Forest	41
4.8	Hilbert Transform and Shannon Envelopes experiment - Regressed Values Error - Bagging	43
4.9	Hilbert Transform and Shannon Envelopes experiment - Regressed Values Error - Random Forest	44
4.10	Shannon Entropy and Energy experiment - Regressed Values Error - Bagging	45
4.11	Shannon Entropy and Energy experiment - Regressed Values Error - Random Forest	45
4.12	LPC options experiment - Regressed Values Error - Bagging	47
4.13	LPC options experiment - Regressed Values Error - Random Forest	47
4.14	LPF and Savitzky-Golay experiment - Regressed Values Error - Bagging	48
4.15	LPF and Savitzky-Golay experiment - Regressed Values Error - Random Forest	48
4.16	Polynomial Fitting experiment - Regressed Values Error - Bagging	50
4.17	Polynomial Fitting experiment - Regressed Values Error - Random Forest	50
4.18	Classification percentage error - Tree Bagging test	54
4.19	Classification percentage error, precision and recall for Normal label, on the polynomial fitting set (101 observations, 61 patients)	55
4.20	Classification percentage error, precision and recall for Normal label, on the LPF filtering set (101 observations, 61 patients)	55
4.21	Influence of Height and Age on the Classification percentage error, precision and recall for Normal labels (101 observations, 61 patients)	56
4.22	Classification percentage error, precision and recall for Normal label, using using different sets of parameters (101 observations, 61 patients)	58
4.23	Linearized Confusion Matrices for the Classification with the 9 clinical regressed parameters	60
4.24	Linearized Confusion Matrices for the Classification with the 9 clinical regressed parameters plus height	61
4.25	Linearized Confusion Matrices for the Classification with the 9 clinical regressed parameters, plus height and age	61

4.26	Linearized Confusion Matrices for the Classification with the 4 most popular clinical regressed parameters, plus height and age, for the Polynomial dataset	62
4.27	Linearized Confusion Matrices for the Classification with the 4 most popular clinical regressed parameters, plus height and age, for the LPF dataset	63
4.28	Clinical parameters' results comparison	64
A.1	Comparing learning algorithms (**** stars represent the best and * star the worst in performance). Reproduction from [3].	74

Abbreviations and Symbols

Medical Related

ATS	American Thoracic Society
BD	Bronchodilator
COPD	Chronic Obstructive Pulmonary Disease
FEF _x	Forced Expiratory Flow at x% of the FVC
FEF _{25%–75%}	Forced Expiratory Flow between 25% and 75%, average flow during the FEM
FEM	Forced Expiratory Maneuver
FEV ₁ /VC	Tiffeneau's index
FEV _t	Forced Expiratory Volume in the first <i>t</i> seconds
FVC	Forced Vital Capacity
MMEF	Maximum Mid-expiratory Flow, currently referred as FEF _{25%–75%}
MMV	Maximum Voluntary Ventilation
PEF	Peak Expiratory Flow
RV	Residual Volume
SVC	Slow Vital Capacity
TD	Tidal Volume
VC	Vital Capacity
V _{TG}	Thoracic Gas Volume
WHO	World Health Organization

Engineering Related

AC	Alternated Current
ADC	Analog to Digital Converter
AM	Amplitude Modulation
ANN	Artificial Neural Network
APF	All Pass Filter
BPF	Band Pass Filter
CART	Classification And Regression Tree
DIY	Do It Yourself
DFT	Discrete Fourier Transform
FFT	Fast Fourier Transform
HPF	High Pass Filter
kNN	k-Nearest Neighbors
LPF	Low Pass Filter
ML	Machine Learning
SL	Supervised Learning
SVM	Support Vector Machines
UL	Unsupervised Learning
USB	Universal Serial Bus
CAGR	Compound Annual Growth Rate

Chapter 1

Introduction

This chapter introduces the dissertation and the work conducted on the project. It's motivation and scope are presented. The problem statement, with the project's objectives and requirements, is described. A brief summary of the business environment is presented. Finally, the structure of the document is detailed.

1.1 Motivation

Asthma is defined by the World Health Organization (WHO)¹ as a chronic disease in which the patients suffer from recurrent events of breathlessness and wheezing. Although the causes of asthma are not fully understood there are risk factors and asthma "triggers" such as allergens, tobacco, and chemical irritants, which can be found with relative ease on an industrialized or developed country. This illness can severely reduce a person's quality of life, if left unmanaged. It is the most common chronic disease among children and currently afflicts 235 million people².

COPD stand for Chronic Obstructive Pulmonary Disease. It covers a number of obstructive diseases that cause limitations in lung airflow³ such as emphysema and chronic bronchitis. The main symptoms of COPD are breathlessness, as asthma, excessive sputum and chronic cough. Left under-diagnosed and untreated, COPD can be a life threatening lung disease. Risk factors include Tobacco smoking, indoor (biomass fuel used for cooking and heating) and outdoor air pollution and occupational dust and chemicals, such as vapors, irritants and fumes. According to the 2004 WHO estimates, 64 million people have COPD, 3 million people died of COPD and it will be the third leading cause of death worldwide in 2030.

mHealth approaches and devices are revolutionizing worldwide health systems and enable to take health practices and low cost equipment to isolated places and to people in need. These platforms are increasingly seen as cost-effective alternatives for hospital care and allow more comfortable monitoring and treatment, regardless their location, and a better quality of life. Although

¹<http://www.who.int/topics/asthma/en/> accessed on 26-06-2014

²<http://www.who.int/features/factfiles/asthma/en/> accessed on 26-06-2014

³<http://www.who.int/respiratory/copd/en/> accessed on 26-06-2014

there has been a significant effort of the medical device industry to make spirometers portable, they are still quite expensive, single purposed and, generally, require external pieces or sensors.

One mHealth platform that has increasingly gained popularity is the smartphone. Smartphones have a considerable computational power for their relatively small size, are easily compatible with several types of devices thanks to serial connections (USB) and wireless communication (Wi-Fi, Bluetooth) and are effortlessly and continuously carried by people.

On the other hand, the transition to smartphones is among the top global mobile networking trends and "the increasing number of wireless devices that are accessing mobile networks worldwide is one of the primary contributors to global mobile traffic growth" [4]. Furthermore, areas in relatively severe poverty have had a booming growth in smartphone use. Africa and Middle East is expected to increase from the current 10% of share of smart devices and connections to 36% until 2018. The mobile data traffic CAGR will also grow to 70% [4]. Currently about 1 in 5 Africans have access to a smartphone [5].

The possibility of reaching such a number of people with less access to medical facilities would certainly improve respiratory healthcare and awareness.

1.2 Problem Statement

1.2.1 Project Objectives

This project aims to develop a mobile and easy to use platform that can help evaluate a patient's respiratory function. This kind of technology can also enable monitoring systems for patients with lung problems such as asthmatics, COPD patients and people with other obstructive and/or restrictive airway diseases. A smartphone application could be a relatively cheap and easy prevention and monitoring system that could enable the early detection of lung illnesses. As guidelines to the development of such application, the platform should have the following properties:

- Self contained,
- Detached from external systems (no external sensors),
- Fast enough (Timely results),
- Precise in the classification,
- Robust to background noise.

The mobile and the self contained characteristics are important in order to reach the most isolated places and avoiding the necessity of constant internet connection for the lung function evaluation and value display. It should be detached from external sensors since requiring disposable and/or replaceable parts both adds to the cost and becomes an additional maintenance problem. The app should also provide timely results, not necessarily in real time. Finally, it should be

precise in the evaluation, since people should not be unduly alarmed of a disease, and should be referred to a physician in the case they do have it or are starting to show symptoms.

In this sense, the methods for the developing of such application, should provide an accurate classification, in a reasonable diversity sound environments, and relatively low delay, i.e., should use the least and the less complex methods possible. Additionally, the system input should consist solely of a smartphone's microphone recordings.

1.2.2 Application Requirements

The basic system should have the following features:

- Audio recording feature,
- Audio envelope generation,
- Appropriate signal filtering options (if necessary),
- Clinical parameter regression,
- Classification of lung function.

This dissertation focuses on finding the best options among these features, in order to provide an implementation design for the aforementioned smartphone app.

1.3 The Company

Center for Health Technology and Services Research (CINTESIS) is a non-profit research and development unit, recognized and supported by the Fundação para a Ciência e a Tecnologia (FCT), with a classification of "Very Good". It is currently hosted by the Faculty of Medicine of the University of Porto (FMUP), located in the recently inaugurated Medical Research Center (CIM) in Porto, Portugal.

Recently, following FCT recommendations, 100 integrated researchers together with 135 PhD students and 139 collaborator members decided to create a large, multidisciplinary and decentralized, translational research unit: CINTESIS, Center for Research in Health Technologies and Services, which will be in effect in 2015. The new unit resulted mostly from merging 3 existing units, classified as "Very Good" in 2007, i.e. the Center of Experimental Morphology, UNIFAI - Unit for Research and Education on Adults and Elders and the former CINTESIS, Center for Research in Health Technologies and Information Systems, it also comprises 50 researchers previously integrated in other R&D units. Altogether, nearly 360 researchers – mostly based in 6 Portuguese higher education institutions, i.e. UPorto, UAveiro, ULisboa, UAlgarve, Polytechnic Institute of Porto (IPP), Nursing School of Porto (ESEP) – are organized into 4 thematic lines, i.e. TL1-Clinical and Health Services Research, TL2-Ageing and Neurosciences Research, TL3-Diagnosis, Disease and Therapeutics Research and TL4-Data and Methods Research, and 16

research groups, addressing various aspects from these 4 main research lines. In addition, many researchers are health professionals working in healthcare institutions (e.g. CH-S. João, IPO-Porto, HCP (CUF-Porto), and CHA-HFaro), which is necessary to implement research projects involving patients.

Such projects include *Digiscope*⁴, which concerns the development of algorithms for diagnosis support, *Chronic Diseases of the Airways*⁵ and *ICAR*⁶, which aim to find the underlying conditions of respiratory diseases, *Smart Diabetes Self-Management Care*⁷, which consists on the development of a mobile application for chronic disease self-treatment, and *CARAT*⁸, which is a mobile application for chronic disease support.

1.4 Structure of the document

This document is structured in five chapters: introduction, literature review, system for classification of lung function, experimental evaluations and conclusion and future work.

Chapter 2 is a review of the literature concerning the relevant medical field, the evolution of spirometry devices and some current projects of eHealth and mHealth. It also includes descriptions of some available solutions related to the work developed.

In Chapter 3 the proposed initial system is described as well as the methods and algorithms that compose it.

The comparison of the several methods, in different settings, is approached in Chapter 4.

Finally, in Chapter 5, the dissertation's conclusions are presented along with some concepts for future work.

⁴cintesis.med.up.pt/index.php/projects/9-projects/49-digiscope-digitally-enhanced-stethoscope-for-clinical-usage

⁵cintesis.med.up.pt/index.php/projects/9-projects/44-chronic-diseases-of-the-airways-contents-and-tools-for-productive-interactions-between-empowered-patients-and-proactive-professionals

⁶cintesis.med.up.pt/index.php/projects/9-projects/56-icar-control-and-burden-of-asthma-and-rhinitis

⁷cintesis.med.up.pt/index.php/projects/9-projects/108-smart-diabetes-self-management-care

⁸www.caratnetwork.org/index.php?lang=pt

Chapter 2

Literature Review

This chapter presents the state of the art of both medical field and technology concerning lung function evaluation. First, a brief description of the medical science behind respiration and spirometry is made. Afterwards, the technology of spirometers is presented, along with a description of their sensors. The chapter will also discuss some eHealth and mHealth projects and the roles of applications (apps) in lung function evaluation. Finally, a brief discussion of some architectural solutions available for smartphone lung function evaluation and spirometry is presented.

2.1 Spirometry

2.1.1 The science of lung function

Respiration is the process by which the body exchanges CO_2 , from pulmonary blood, for O_2 from the atmosphere, in order to meet the body's metabolic demands. Physiologically, the process consists on the integration of the ventilation, blood flow and gaseous diffusion.

Spirometry is a widely used, non-invasive test of ventilatory response and assessment of the mechanical or bellow characteristics of the patient's pulmonary system. Along with arterial gas tension or pulmonary gas exchange measurements it enables an overall assessment of lung function which can be used for detection, differentiation and diagnosis of several respiratory diseases. It is also useful for tracking a respiratory disease progression or improvement and therapeutic response over time. The test itself consists of a timed measurement of the dynamic or respired lung volumes and capacities during forced expiration and inspiration, enabling the quantification of how quickly and effectively the lungs are depleted and filled [1].

However, the lungs are never completely emptied since that would cause the lungs to collapse. The remaining gas volume is called the residual volume (RV). During the normal respiration cycle, the lungs keep within a volume range called the Tidal Volume (TV). The forced expiration maneuver (FEM), that is performed during spirometry tests, consists of instantaneously inspiring and completely filling the lungs and then forcefully expiring, emptying the lungs. The lung volume variation during one of these tests is depicted on Figure 2.1. The maximal slow excursion of volume after a maximum inhalation is named the Vital Capacity (VC).

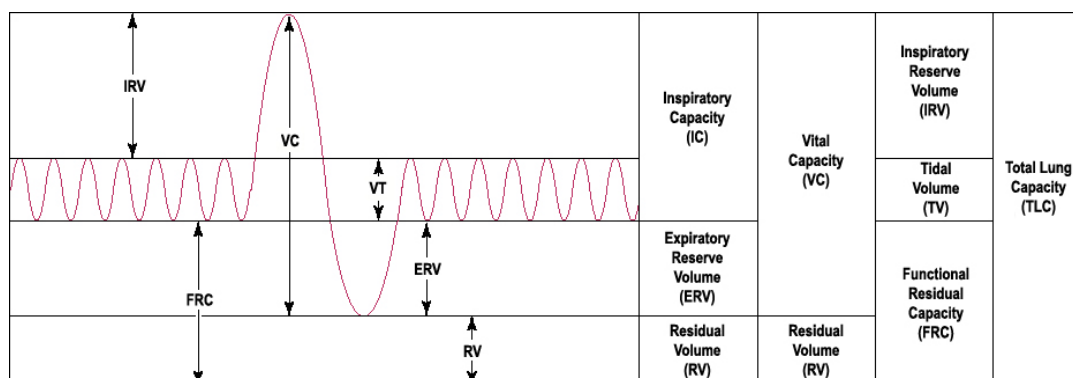


Figure 2.1: Respiratory Tidal Volume and Forced exhalation

In the 1800s, Hutchinson developed a water-sealed spirometer that enabled the measurement of VC. He later observed that the VC was proportional to a person's height and created tables to estimate a healthy patient's expected VC. The Forced Vital Capacity (FVC) is an enhancement of the simpler VC test.

During the 1930s, Barach observed that healthy patients exhale more quickly than patients with asthma or emphysema and noted that the airflow was a key aspect of detecting airway obstruction.

In 1947, Tiffeneau described the FEV_1/VC , ratio between the Forced Exhaled Volume under one second and the Volume Capacity, as an index of obstruction, hence the name Tiffeneau index. However, the medical community has embraced the FEV_1/FVC index, rather than Tiffeneau's, partly because it implied another, slower, exam and the results evaluate almost the same lung characteristics.

Around 1950, Gaensler observed that, for healthy individuals, approximately 80% of the FVC was expired in the first second and the FVC in 3 seconds. This established a baseline for comparison between obstructed and unobstructed airways.

In 1955, Leuallen and Fowler introduced the Maximum Midexpiratory Flow Rate (MMFR) which is calculated by $1/2FVC / (t_{75\%} - t_{25\%})$. It can also be named just Maximum Midexpiratory Flow (MMEF or MEF). This provides an approximate, somewhat graphical way, to quantify the center slope of the Flow-Volume exhalation curve. This characteristic of the curve's shape is another method used by physicians to characterize lung function. Later the term was standardized as Forced Expiratory Flow 25%-75% ($FEF_{25\%-75\%}$). However, for shortness of writing, the terminology MMEF will be used on Chapter 4.

In the 1960s, Wright and McKerrow spread the concept of monitoring asthmatic patients using the Peak Expiratory Flow rate (PEF or PEFr). PEF tests, referred as bronchial challenge or bronchial provocation tests, consisted of performing forced expirations before and after the inhalation of bronchodilator. The analysis was based on the percentage of change [2]. The PEF also became important for a simpler tracking of intra-daily specific change patterns in respiratory function, such as those that occur for general and occupational asthmatics [1].

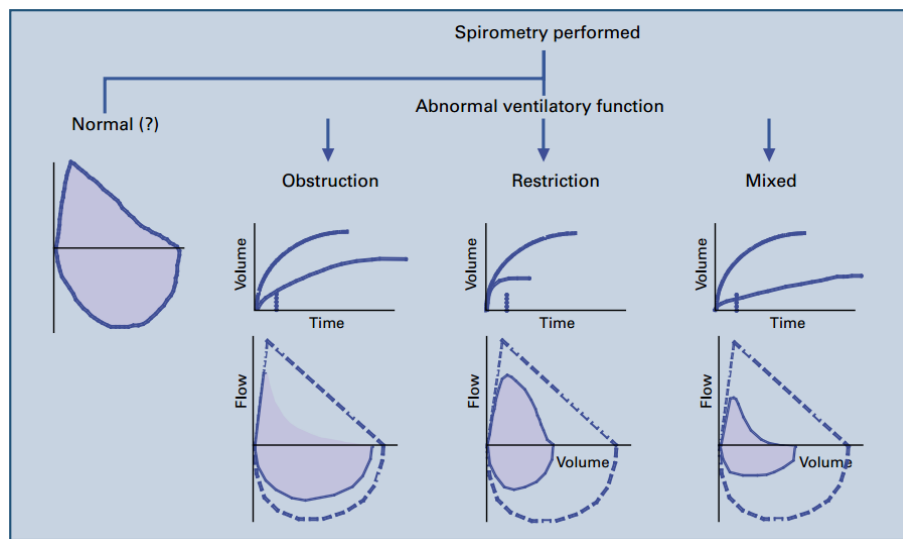


Figure 2.2: Types of ventilatory defect with typical spiromograms and flow-volume curves. Image taken from [1]

There are also methods to analyze lung function, less focused on forced mechanic characteristics and more on gas exchange. Examples of those tests are carbon monoxide diffusion capacity and blood oxygen and carbon dioxide levels measurements [1]. However, "Spirometry is recommended as the "Gold Standard" for diagnosis of obstructive lung disease by the National Lung Health Education Programme (NLHEP), the National Heart, Lung and Blood Institute (NHLBI), the World Health Organization (WHO), and numerous other organizations concerned with diagnosis of lung diseases" [2], and thus, the focus of the dissertation.

2.1.2 Lung function parameters

Although there are other possible clinical parameters that can describe lung function, only those on Table 2.1 will be measured, since those are the most popular. However, $FEF_{25\%-75\%}$ and FEF_x are seldom considered useful on the lung function evaluation since they are highly dependable of the patients' cooperation. Cust1 is a custom parameter proposed by [6] that, allegedly, provided values more stable than FEV_1 . It is calculated as $FEF_{25\%-75\%} / FEV_1 \times 100$.

The primary components refer to the wider airways and the secondary refer to the smaller airways. Generally, obstruction affects the rate at which the smaller airways can expel air, and can be noticed with the decrease of FEV_1 and, consequently, FEV_1/FVC . $FEF_{25\%-75\%}$ may be indicative of obstruction of medium and small airways, though it should not be used alone for diagnosis [2].

Lung restriction is characterized by the strong reduction of total exhaled air, as well as a relatively slight decrease of airflow exhalation rate. This can be identified by a slight reduced

¹A physician may also evaluate the patient's collaboration based on the PEF performance relative to the reference values, which are calculated from the patient's vital characteristics (height, age, ect.).

Table 2.1: Clinical parameters used

Parameter	Type	Meaning/Airway component measured
Forced Vital Capacity (FVC)	Measured	Primary and Secondary (All)
Forced Expiratory Volume under 1 second (FEV ₁)	Measured	Primary
FEV ₁ /FVC index	Calculated	Ratio of Primary over All
Peak Expiratory Flow (PEF)	Measured	Primary and patient collaboration ¹
Forced Expiratory Flow 25% – 75% (FEF _{25%-75%})	Measured	Secondary
Custom parameter (Cust1)	Calculated	Secondary over Primary
Forced Expiratory Flow at 25% (FEF _{25%})	Measured	Secondary
Forced Expiratory Flow at 50% (FEF _{50%})	Measured	Secondary
Forced Expiratory Flow at 75% (FEF _{75%})	Measured	Secondary

FEV₁, a strongly reduced FVC and, as a consequence, a slight increase of the aforementioned index.

Mixed Abnormalities are rare conditions in which both obstruction and restriction features manifest. It is characterized by overall reduction of VC, FEV₁ and Tiffeneau index [7]. This type of classification shall not be analyzed in detail due to its rarity and the fact that no reference values are used, either for the FEV₁% predicted or other similar typical parameters.

Examples of the respective lung functions can be observed on Figure 2.2.

2.2 Spirometry Technology

The precursor of the modern spirometer was introduced by Hutchinson around 1844. It was a volume displacement water-sealed device. In recent years the sensor paradigm has shifted towards the flow measurement type thanks to the evolution and wide-spreading of electronic devices. Also, it became possible to measure lung function parameters, using several other methods.

During the course of lung function testing history many types of systems have been developed. Some of them enable exams that complement each other and others are specific for providing data from particular respiratory related diseases. A short presentation of each medically accepted spirometry device, and respective sensors, is provided, as described on [2].

2.2.1 Pneumotachometers

From Greek, *pneuma*, "wind" or "breath", *thacos*, "speed" and *metron*, "measure", pneumotachometer means that which measures breath speed. "The flow sensors that historically have been, and continue to be, the mainstay of the respiratory laboratory utilize flow resistors with approximately linear pressure-flow relationships. These devices are usually referred to as pneumotachometers" [8].

2.2.1.1 Volume Displacement Spirometers

As the name implies, these spirometers measure directly the shifting fluid volumes. Since the designs of this kind of devices were made during a time prior to the electronic massive growth they are mainly based on mechanical actuation. The movement caused by the breathing expirations is transferred to a stylus which could write on a moving chart. More recently, with the development of computers, potentiometers were incorporated so the signal could be better analyzed.

Water-Seal Spirometer: This device consists of a large bell suspended in a container with its open end submerged. Breathing into the spirometer moves upwards the bell by an amount proportional to the volume expired. During an inspiration the bell submerges beyond the stable level. For many years the bell had a pen attached, that moved sympathetically with the bell. The pen recorded on a rolling chart, the kymograph, the respective volume. In recent years the bell activates a potentiometer handle that varies a DC voltage output. This signal could later be digitized with an ADC and processed by a computer.

Dry Rolling-Seal Spirometer: This spirometer is cylindrically shaped holding a lightweight piston inside. A flexible plastic seal couples the piston to the cylinder wall. The seal rolls on itself when the piston moves slightly with the gas' volume displacement. The piston shaft rides on a linear bearing which actuates a rotary potentiometer. This generates an analog signal for flow and volume that can be fed to an ADC and analyzed.

Bellows-Type Spirometer: This kind of device uses a collapsible bellow that folds or unfolds depending on the fluid's flow direction. The common bellow design resembles a flexible accordion. One end is fixed and the other moves freely with the expirations. This movement actuates either a pen that writes on a moving chart or a potentiometer similar to the other volume displacement spirometers.

2.2.1.2 Flow-Sensing Spirometers

The following spirometers measure air flow based on several different principles. All of them include the respective sensor and a tube to contain the expiration. Figure 2.3 illustrates the following sensor types.

Turbine: The turbine, or respirometer, is an instrument that "consists of a vane connected to precision gears". The vane's rotation is triggered by the gas flowing through it. This sensor is more suited to measure VC. The sensor's configuration can also use a light source and a photoreceptor to retrieve the number of rotations per unit of time, just like an odometer for motor controllers.

Pressure Differential Flow Sensors: This types of sensors make a pressure differential measurement. The sampling openings are placed one before and another after a resistive element, in

the middle of the sensor tube. The element accelerates the gas by narrowing the gas flow opening, while reducing the pressure. This is called the Venturi effect. While changing the gas' velocity and pressure, the element also ensures laminar flow.

The resistive element can either be a bundle of capillary tubes, the Fleisch type, or a mesh screen or membrane, the Silverman or Lilly type. The Fleisch type is more reliable than the Lilly type. Though the Lilly type is better suited for measuring widely varying flows. The disadvantages of these kinds of measurement devices are that they are very sensitive to local atmospheric pressure, temperature and humidity. As a consequence, they need to be calibrated daily.

Heated-Wire Flow Sensors: The heated wire method is based on the cooling effect of a flowing gas. The wire is preheated to a fixed amount and, during the expiration, the gas' flow reduces the wire's temperature. The instantaneous energy consumption to reheat the wire gives a flow value at that time.

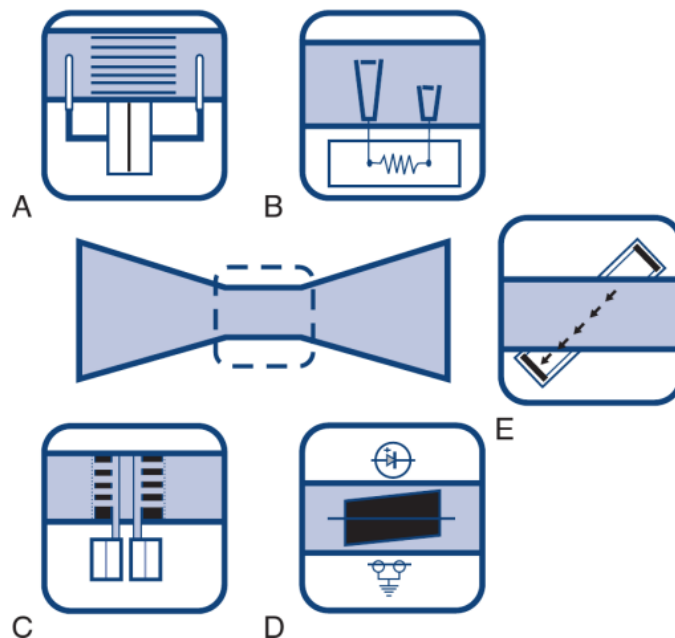


Figure 2.3: **Common Flow Sensing Devices (Pneumotachometers).** A-Fleisch type sensor, B-Heated wire sensor, C-Pitot tube sensor, D-Turbine sensor, E-Ultrasonic sensor. Image taken from [2].

Pitot Tube Flow Sensors: The Pitot tube sensors are based on the principle that states that the pressure of a fluid is related to it's density and velocity. In general, a spirometry Pitot tube sensor has, at least, two sets of small tubes, one facing each flow direction for measuring both expiration and inspiration. The tubes are connected to pressure transducers.

Ultrasonic Flow Sensors: A sound wave is a compression wave that is influenced by the means through it travels. The flowing of gas can increase or decrease a sound's speed depending of the

relative direction it travels. The sensor consists of two oppositely faced ultrasound transducers, angled with the tube. By measuring the sound's traveling time the flow measurement can be integrated into volume. There is also a different type of gas flow measurement that can be obtained by quantifying the ultrasound's frequency shift, based on the Doppler effect.

The design includes a disposable tube which is introduced between the sensors. This allows a great advantage over several other spirometer types since it needs no calibration and solves the problem of patient cross-contamination.

2.2.2 Peak Flow Meters

This instrument was designed to measure Peak Expiratory Flow (PEF). It provides a simple measurement that aids asthma monitoring with an inexpensive device. The meters themselves tend to be made disposable. These meters work by forcing air through a resistor or a flow tube with a movable indicator.

2.2.3 Plethysmographs

From Greek, *plethysmos*, "increase" and *graphein*, "to write". The plethysmograph is an instrument that measures the change in volume within an organ or body.

There are two types of body plethysmographs: the constant-volume variable-pressure and the flow or variable volume plethysmographs, **pressure plethysmograph** and **flow plethysmograph**, respectively. Both are devices that include a sealed box which the patient has to be in. They are used to measure the thoracic gas volume (V_{TG}), the specific airway resistance (sRaw) and its derivatives.

2.2.3.1 Pressure Plethysmograph

This spirometer is based on an adaptation of Boyle's law. The change in volume inside the chamber is inversely proportional to the pressure variation. The pressure change is caused by compressing and decompressing the atmosphere inside the box, including the patient's chest. If the temperature remains constant, the pressure change corresponds to a specific volume change. Since the chamber has rigid walls and is sealed, its free volume experiences the same, mirror-image shift volume as the lungs [9].

2.2.3.2 Flow Plethysmograph

This kind of plethysmograph uses a flow transducer in the chamber's wall. This device measures the volume changes inside the box. As gas is compressed and decompressed, flow passes through the chamber's opening. That flow is integrated and volume change is recorded as the sum of the volume compressed and the one that left through the opening. With this device there are also some other tests that can be conducted using a pneumotachometer connected to the outside room.

2.3 eHealth and mHealth approaches

The rise of the Internet's generalized access and widespread on the mid-1990s enabled the emergence of a frontier field, where web based services fused with health related systems [10]. World Health Organization defines eHealth as "the transfer of health resources and health care by electronic means" and encompasses the delivery of health information by the Internet and other telecommunication means, the use of Information Technology and e-commerce to improve public health services and the use of e-commerce and e-business practices in health system management [11]. Some common examples of eHealth are telemedicine and remote patient monitoring.

As mentioned on chapter 1, in recent years, there has been a miniaturization of electronics as well as an improvement of mobile devices' computing capability. The rapid penetration of technologies such as smartphones, personal digital assistants (PDA), tablet computers, wireless wearable bio-sensors and disease monitoring devices has enabled a significant shift of medical decision-making tasks from the medical community to the patient side. Such tasks range from diagnosis, selection of appropriate treatment and prognosis. This change in perspective has the focus of improving the awareness of a patient's own health condition and consequent empowerment [10]. Mobile health, or mHealth, although not defined as a standard yet, is a field of eHealth that makes use of the aforementioned types of devices in order to improve medical and public health [12].

2.3.1 Games and Applications

Throughout spirometry history there have been several technological advances. As described on Section 2.2, sensory methods have shifted from purely mechanical to partially electronic. As a consequence, spirometry apparatuses have become computerized and became able to include secondary functions or add-ons. Computers have maintained the tendency of getting smaller, following a corollary of Moore's law². This means that there has been a reduction of the size of the platforms where spirometry related software run. In recent years, significant effort has been focused around smartphones since they are continuously carried around and include a significant amount of electronic devices.

Games for compliance: The forced expiration maneuver is somewhat difficult to execute correctly. This case is specially true for children that need to make a lung function evaluation. Therefore, pulmonary function laboratories commonly have computer programs that guide children through the process, most of the time without them even noticing. Usual tests comprise computer games that are played using breathing apparatuses, such as breathing forcefully through a tube. This enables a better cooperation from the patient since kids rarely wish to lose a game.

²Moore's law is an empirical prediction which states that within roughly 18 months a chip performance will double, due to the miniaturization of transistor technology.

On the other hand, this gaming environment has also been reaching the long term vital parameter monitoring. Azmo the Dragon³, a Microsoft Imagine Cup finalist, is a game developed by students of Rice University in Houston. This game's aim is to get children excited on tracking their own lung function and better manage their condition. The child temporarily forgets about the illness and becomes a player, more precisely, a dragon that attacks castles, by breathing fireballs, in order to retrieve its siblings and defeat the evil empire.

Another Imagine Cup entry that focuses on lung function monitoring is Ki-Breath⁴. This platform combines a series of mini-games, which acts both as a complement to physiotherapy and encouragement to patients with respiratory problems, such as suffering from cystic fibrosis, to fight against the disease.

Finally, mCOPD, which will be discussed on Section 2.3.2, was designed to include lung function monitoring games [13].

Games for entertainment: With the growth in smartphone use and a rising concern about respiratory diseases, several independent application developers have created sets of games that relate to spirometry. Many measure the loudness recorded by the smartphone's microphone to actuate something. Some of them actually advertise spirometry, though they appear under the game application category.

An example of an entertainment oriented application is Spirometer by Patrick Giudicelli⁵. The purpose of this application is to raise a ball above a red line, by expiring to the smartphone's microphone, in order to score how long an exhalation takes.

Another game, somewhat related to lung function evaluation, is Blow Up Balloon⁶, from MTM HZSTUDIO. The objective is to take a deep breath and forcefully exhale to the microphone. The score is proportional to how fast the balloon explodes, which might be based on exhalation time and loudness.

Mobile health applications: There is a significant amount of developers that create smartphone applications that measure spirometric parameters. Those applications usually have an external sensor that is connected by Bluetooth, Wi-Fi or USB to the smartphone. The application becomes a mean to relay data, observe it and update a patient's profile. Some solutions are also compatible with multiple sensors for different health problems such as heart rate monitors, pulse glucose and cholesterol meters, oximeters and so on.

SpiroID⁷, from Thor Laboratories, is one of such applications. By using the external SpiroTube Mobile Edition sensor⁸, which connects by Bluetooth to the smartphone, real, medical grade, spirometry curves and clinical parameters can be measured.

³<http://www.bethherlin.com/167918/2433983/projects/azmo-the-dragon>, last accessed on 20-06-2014

⁴[https://www.imaginecup.com/finalists/team?teamName=ki\(breadth\)](https://www.imaginecup.com/finalists/team?teamName=ki(breadth)), last accessed on 20-06-2014

⁵App store link: <https://itunes.apple.com/us/app/spirometer/id315550839>, last accessed on 20-06-2014

⁶App store link: <https://play.google.com/store/apps/details?id=com.hz.game.balloon>, last accessed on 20-06-2014

⁷Web store page: <http://play.google.com/store/apps/details?id=com.thorlabor.spiroid>, last accessed on 20-06-2014

⁸Thor Laboratories SpiroTube page: <http://thorlabor.com/en/products/pulmonary-medical-devices/spirotube-mobile-edition-bluetooth-spirometer/> last accessed on 20-06-2014

In addition to this SpiroID, all prototypes that are discussed on the following section constitute a form of mHealth applications, despite not needing any external sensor (embedded microphone only).

2.3.2 Prototypes and developing projects

There has been a significant effort to develop spirometric devices focused on the availability of materials or cost efficiency. This aims to reach areas with less access to health care and in isolation, as described on Section 1.2.

A team from Penn State University developed a Lilly-Pneumotach type spirometer from easily accessible materials [14]. This Mashavu project prototype was designed as a do it yourself (DIY) concept enabling the wide-spreading of the design. It used a pressure differential electronic sensor to measure the airflow and was made with PVC tubing, coffee stirrers, protoboard as a mesh along with cotton cloth for filtering.

Another relevant project came from a multidisciplinary team of students from the University of Washington in St.Louis. They worked towards a low cost spirometer [15]. Their solution consisted of using a microphone as a pressure transducer and coupling it to a custom designed breathing piece. The piece was relatively cheap to produce, could be easily sterilized and was of robust manufacturing.

Smartphones have also been a research focus. A group from Radbound University [16, 17] developed a system in which measurements were made with external sensors (a pulse-oximeter and a custom spirometer), which were directly connected to a bluetooth sensor interface that relayed the data to a smartphone. This smartphone app also integrated a questionnaire, communication with a web-center and computed the predictions. The multiple COPD exacerbation classification were made by a Bayesian Network constructed with medical help.

TeleSpiro [18] is a combination of custom spirometer design and interface for computer, tablet or smartphone. The spirometer consists of PVC tubing and a differential pressure sensor which is mounted on a flash drive like circuit board. This board also contains a digital humidity/temperature sensor and can communicate using a male USB 2.0 port.

One project that motivated this dissertation was the *SpiroSmart* application [19] addressing the Ubiquitous Computing field. The application consists of a somewhat complex software that records the patient's breath while giving incentive and showing the performance. It employs signal processing and machine learning methods to measure lung function.

Two ensuing projects, that started simultaneous and independently, as MSc theses, were the *mCOPD* [13] and the *SpiroApp* [6]. Both systems use forced expiration sounds, recorded with the built-in smartphone microphones, analyze them and provide some means of clinical information concerning lung function, similarly to *SpiroSmart*. *mCOPD* is focused on the interactivity of Exergaming (fusion between Exercise and Gaming), while *SpiroApp* tries to provide a more robust implementation of lung function evaluation using the microphone.

2.4 Software Solution Review

This section describes in detail the systems' architecture of some relevant projects that tackle similar problems as this dissertation project. Table 2.2 presents a brief summary of the aforementioned projects.

2.4.1 SpiroSmart

This system's goal was to accurately replicate the spirometer's Flow-Volume and Volume-Time curves, as well as to obtain several clinical parameters from those curves.

The Signal Processing stage first converted the audio input into pressure at the lips and airflow at the lips. Then all those signals were fed to envelope generating models, namely Hilbert Transform, Linear Predictive Coding (LPC) and Mean of Resonances. All of the envelopes were fed to low pass filters to remove noise and finally, smoothed with Savitzky-Golay polynomial filter. All of these methods are described in greater detail in Section 3.2.

On the Machine Learning end, the smoothed envelopes were extracted for parameters (PEF, FEV₁ and FVC) which were regressed with bagged decision trees. Those values were later clustered with k-means, which predicted their final values. These values were used to normalize the several envelopes and, combined with raw envelopes, the final envelopes were obtained, using Bagged Regression Trees and Conditional Random Fields (CRF) on 15ms frames. These last curves were filtered for abnormal events such as non-monotonic Volume-Time curves and, finally, clustered with k-means, producing the final curves.

2.4.2 SpiroApp

The SpiroApp's implementation records the audio from the FEM, calculates FFT magnitudes on 50% overlapping, 100ms frames and regresses the Flow-Time curve with a fourth order polynomial function. It's main objective is to provide a simple way to calculate an approximation of a patient's vital parameters, mainly for classification in "good", "maybe bad" and "bad" lung function. The use of just polynomial regression aims to provide a more robust system against ambient noise and to reduce the implementation space and speed requirements.

2.4.3 mCOPD

This system records the sounds with the microphone and converts the air pressure to air velocity based on the microphone response model. The pre-processed signal is low pass filtered to remove noise and the frequency components are obtained with a 64-point FFT. This result is regressed either by Nearest Neighbor Matching or by a second order polynomial, depending of the platform requirements. From the regressed Flow-Time curve the Volume-Time is obtained and the clinical parameters FEV₁, FVC, and FEV₁/FVC are calculated. These are displayed, as well as the total duration of the exhalation.

System	Shape regression ⁹	Classification / Diagnosis	Clinical Parameters	Project Stage	Objective
SpiroSmart [19]	Yes	No	Yes	Prototype	Curve regression and clinically accepted parameters extraction
SpiroApp [6]	No	Yes	Yes	MSc Dissertation / Prototype	Gross evaluation as healthy or not.
mCOPD [13]	No	No	Yes	MSc Dissertation / Prototype	COPD detection, evaluation and monitoring based on FEV1 and FVC.

Table 2.2: Microphone based smartphone solutions' characteristics

⁹By shape regression it is intended as accurately representing the shape of the exhalation, in a similar fashion as in a spirometer report.

Chapter 3

System for Lung Function Classification

This chapter presents the classification system and methods that were tested in the course of this project. First, the testing architecture is described as well as the reasons for comparing the several sets of methods. Then, the methods and algorithms are explained.

3.1 Initial System Architecture

3.1.1 Concept and Structure

The project's objective is to use microphone recordings of people performing the forced expiratory maneuver (FEM) in order to classify the respective lung function. The system is structured in two main stages: the signal processing stage and the machine learning stage.

On the signal processing stage the recordings' envelopes are calculated in order to extract those curves' clinical parameters, just as it would be done on a spirogram. Initially, the recordings undergo some pre-processing conversion models, namely the Inverse Radiation and the Flow Conversion. Then, both the original audio and the other steps are fed to the envelope generating functions: Hilbert Transform, Shannon Entropy and Energy, LPC and Mean of Resonances (MRess). Several envelope functions are used since they have different characteristics and provide different aspects of the sounds to be analyzed. The envelopes are then subject to an LPF¹ or smoothed with a moving average (window size of 4% of the sampling frequency). Finally, these results are fed to a Savitzky-Golay smoothing filter (order 3 and size 11). The unprocessed envelopes are also subject to a polynomial approximation (4th order). The clinical parameters are extracted from both the polynomial fitting and smoothing results. Figure 3.1 presents the initial signal processing diagram, containing all the methods evaluated.

The machine learning stage can be divided in two other stages: the regression stage and the classification stage. The total of processed envelopes is 54 resulting in 486 parameters extracted. These are far too many comparing with the dataset available, hence the parameter reduction with the regression stage. Each clinical parameter from each recording is regressed, resulting in the 9

¹Low Pass Butterworth filters, for Hilbert and Shannon envelopes: 2nd order at 50Hz followed by 1st order at 100Hz; for LPC and MRess: 1st order at 2560Hz.

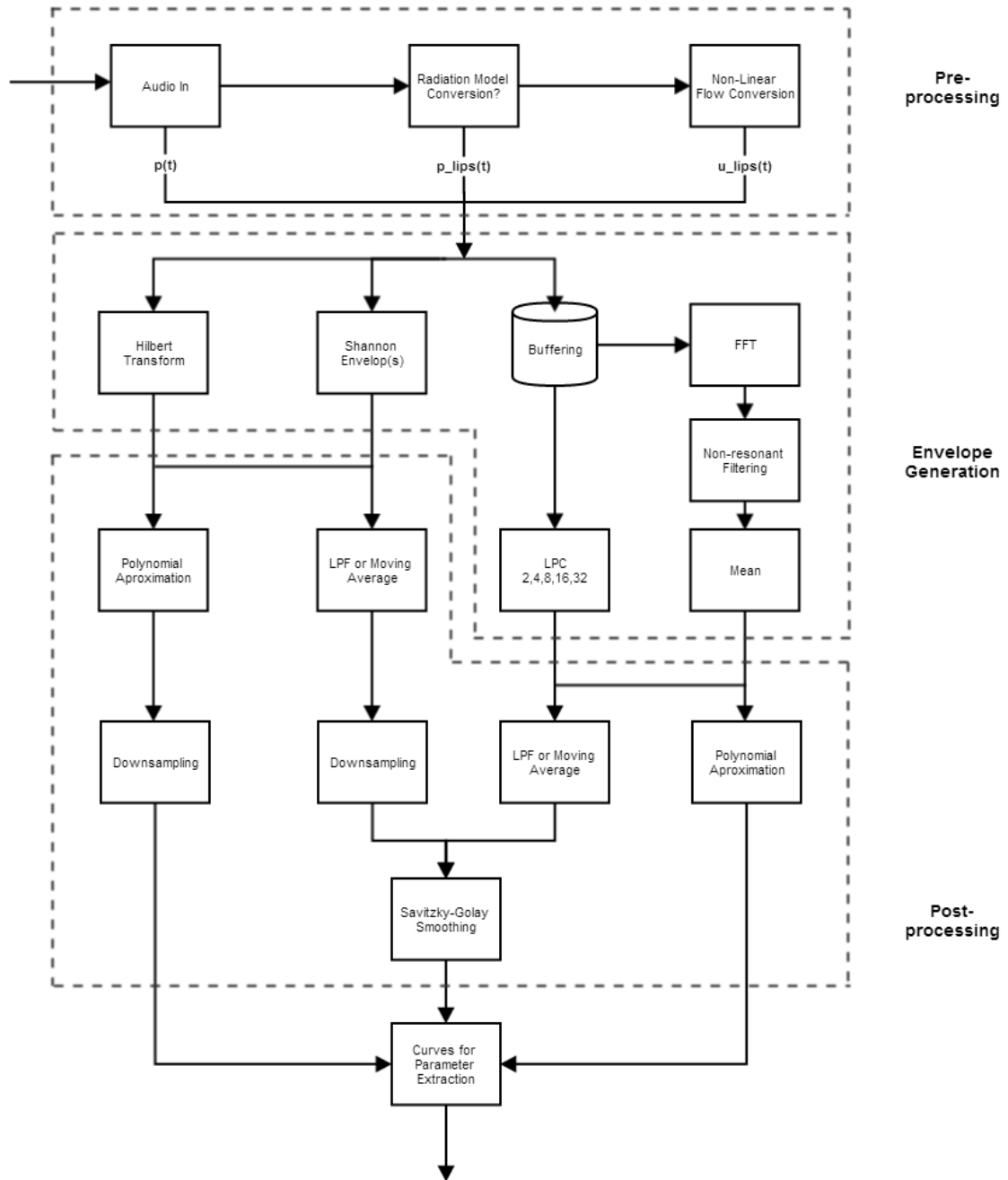


Figure 3.1: Initial Signal Processing stage Diagram

parameters described on Section 2.1.2. The regression models used were Regression Trees employing the following ensemble methods: Bootstrap Aggregation (Bagging) and Random Forest. These parameters, along with some patient's vital information (height and age), were fed to the classification stage. At this point, there were still quite some parameters/dimensions where to scatter the classifications, therefore linear classifiers such as Classification Trees and SVMs were tested. It was also tested the Naïve Bayes classifier.

3.2 Signal Processing

This section discusses the several signal processing methods tested on the project. First, some audio cropping and conversions are presented, then the envelope generating functions are explained and finally, the filtering and smoothing methods are detailed. A discussion concerning the choosing of the implemented methods is presented on Section A.1.

3.2.1 Pre-processing

Automatic Crop Algorithm: The recording environment varies significantly from session to session. Additionally, each recording contains the rapid inspiration sound, which is not intended to be explored, and occasional background noise, such as people talking or machine sounds. In order to cope with these, a method to automatically segment the recordings was devised.

Clinical spirometers, although do not have to worry about flow direction, usually provide a method to calculate the zero time instant, named Back Extrapolation. This value defines the start for all timed measurements since it corresponds to the start of the forced expiration. It changes slightly the time at which the patient starts the expiration but provides a mean to discern if the FEM had a hesitant start or if the effort was maximal [20].

The back extrapolation method was adapted to find the approximate time of the beginning of the exhalation. First, the LPC envelope of the recording is obtained, as described on Section 3.2.5. Then the envelope is subtracted of an arbitrary ratio of the maximum flow value (9% of PEF) in order to retrieve the two zero-crossings immediately before the PEF. By removing the part of the envelope prior to the average of those instants, the most relevant part of the inspiration is discarded. The back extrapolation itself consists of finding the zero-crossing of the steepest slope of the Volume-Time curve [20]. This slope coincides with the instant when occurs the PEF. The slope is calculated using the two neighboring points of the PEF in the Volume-Time curve, which corresponds to a, roughly, 47 ms period. The recording crop is made around 80 ms before the zero time to ensure a margin for the envelope calculation algorithms to work with. This back extrapolation and cropping process is shown in Figure 3.2 and Figure 3.3 is an enlargement of Stage 3.

A second cropping method is used for the ending of the recording. The method consists of a sliding window, starting at the absolute maximum of the recording, which moves to the end of the recording, in order to reach noise level. This noise level is defined as an arbitrary ratio (2%) of the absolute maximum value of the recording. This way, the noise after the FEM is removed from

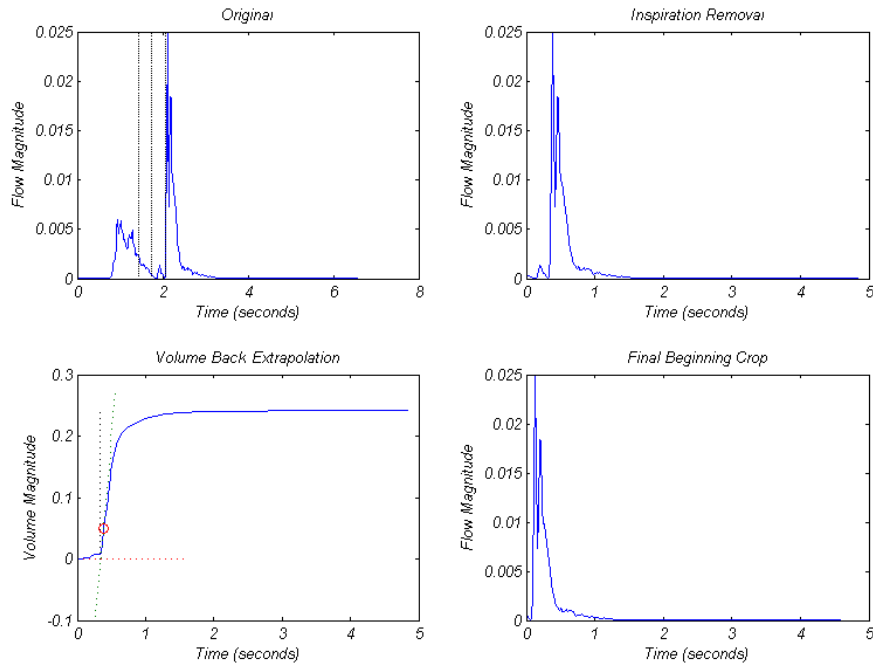


Figure 3.2: Zero time Back Extrapolation algorithm. Top left: Stage 1, original Flow-Time envelope, grossly separating the inspiration from the expiration portions. Top right: Stage 2, Flow-Time Envelope with the inspiration portion removed. Bottom left: Stage 3, Volume-Time envelope back extrapolating the zero time. Bottom right: Stage 4, Flow-Time envelope with pre-zero time values removed.

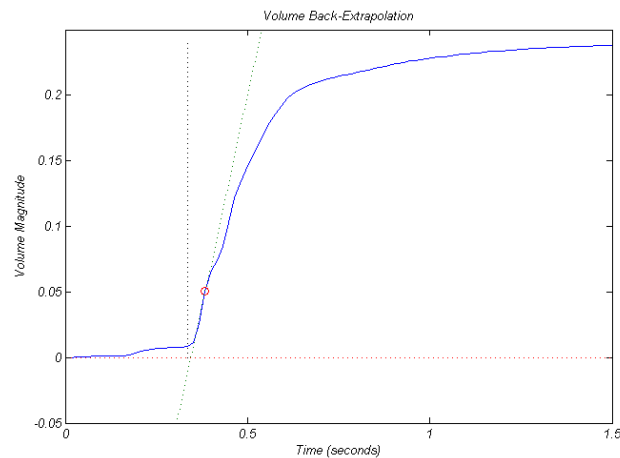


Figure 3.3: Stage 3 of the Zero time Back Extrapolation algorithm (enlargement). The zero time is where the PEF time tangential line intercepts with the zero volume. The PEF time is marked as a circle.

analysis. The window size is also arbitrary (used 5% of the recording length, moving with 25% of overlap). The results of the two methods are shown on Figure 3.4.

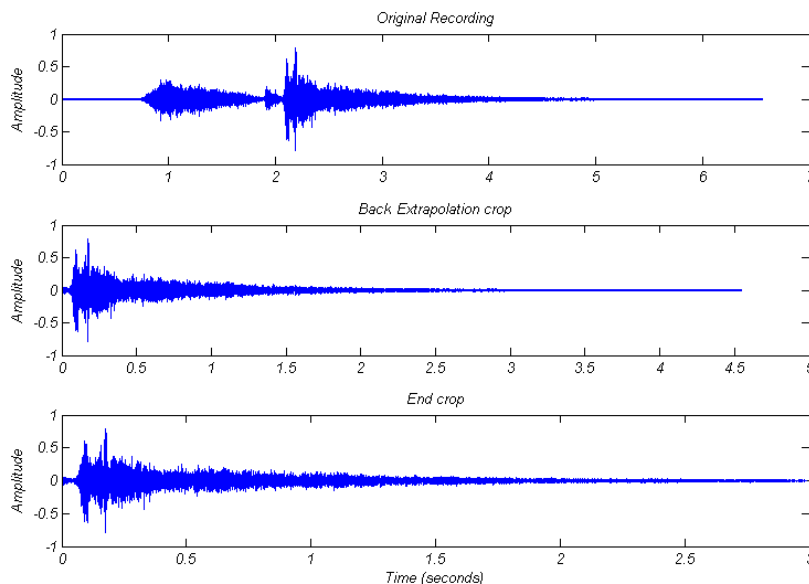


Figure 3.4: Automatic crop steps. Top image: Original recording. Middle image: Recording after crop from the back extrapolation algorithm. Bottom image: Recording after the previous crop and the final, moving window crop.

Inverse Radiation Model: The input of the classification system is a microphone recording of the FEM. However, a microphone can only measure the air pressure around it. Spirometers measure the airflow that comes from the mouth. Therefore, it seems reasonable to approximate the airflow at the mouth based on conversion models [19].

The transfer function that translates the pressure loss around the patient's body and across the distance between the mouth and the microphone is given by equation 3.1. It constitutes an approximation of the breath as a spherical baffle in an infinite plane [21]. D_{arm} is the length of the patient's arm/distance to the microphone, C_{head} is the head circumference (approximated from the subject's height, just like D_{arm} , and c is the speed of sound. D_{arm} is calculated proportionally to the height and according the gender, using values from [22]. C_{head} values are estimated based on the 50% centiles of [23], also according height and gender.

$$H(e^{j\omega}) = \frac{P(e^{j\omega})}{P_{lips}(e^{j\omega})} \sim \frac{j\omega C_{head}}{D_{arm}} \exp\left(-\frac{j\omega D_{arm}}{c}\right) \quad (3.1)$$

The intended model is the inverse of this transfer function. However, since it is non-causal, the inverse is converted to the time domain, h_{inv} , for a "manual" convolution of the delayed time

series, given by equation 3.2.

$$h_{inv}(t) = \frac{D_{arm}}{C_{head}} u\left(t + \frac{D_{arm}}{c}\right) \quad (3.2)$$

On the other hand, the operation involves a step function with a simple delay, which does not have any relevant meaning, since the recordings are first fed to an autocrop algorithm. This way, the model was reduced to a gain whose importance will be discussed on chapter 4.

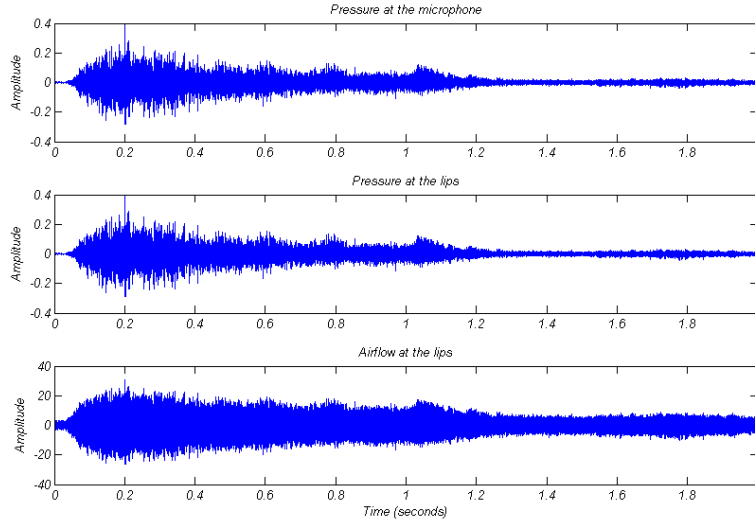


Figure 3.5: Pre-processing conversion models' results

Pressure to Flow Conversion Model: The non-linear equation 3.3 models the pressure drop across the lips ($p_{lips}(t)$) to flow rate through the lips ($u_{lips}(t)$) considering the turbulent airflow [21]. r_{lips} is the radius of the mouth opening. To the best of our knowledge, there are no studies concerning adult mouth open radius and, therefore, these values had to be estimated based on a linear regression produced with values from an online proportion calculator². In turn, these values were calculated based on the height and gender.

$$u_{lips}(t) \sim 2\pi r_{lips}^2 \sqrt{2p_{lips}(t)} \quad (3.3)$$

It should be noted that the expressions for both the inverse radiation and pressure to flow conversion models are only proportional. This is due to the removal of scaling constants, that are not meaningful for the intended analysis since the microphone is not calibrated and may have a non-linear model. Also, only relative values are relevant for the regression stage. The results from the conversion models are depicted on Figure 3.5, as they would have been calculated on

²<http://hpc.anatomy4sculptors.com/>

the system, i.e., the results from the Inverse Radiation model become the input of the Airflow conversion model.

3.2.2 Mean of Resonances

When a system's response has a tendency to oscillate at a specific range of frequencies it is said to resonate. As pointed out in [19], it is reasonable to assume that airflow should be proportional to the magnitude of frequencies resonating on the vocal tract and mouth opening. A spectrogram of a recording is made: the signal is divided into overlapping chunks (of 31.25ms) from which the frequency components are obtained by an FFT.

The *Fast Fourier Transform* (FFT) is an algorithm commonly used in sound processing and analysis. It is an efficient way to compute the *Discrete Fourier Transform* (DFT) and its inverse. This process can be viewed as the transformation of an amplitude and time signal into its frequency power spectrum contents.

Let x_0, \dots, x_{N-1} be complex numbers:

$$\text{DFT: } X_k = \sum_{n=0}^{N-1} x_n e^{-i2\pi k \frac{n}{N}} \quad k = 0, \dots, N-1. \quad (3.4)$$

where n stands for a discretized instant of time, k is a discretized frequency, x_n is the n th complex sample from a N th periodical signal and X_k is the k th value of the frequency spectrum. After the frequency images are sequenced and the spectrogram is attained, only the resonant frequencies are saved. Every value lower than 20% of the global maximum is filtered as well as any resonance less than 125ms. Figure 3.6 represents the stages of filtering the spectrogram. The mean of all remaining frequency contents is obtained, culminating in an envelope.

The conservation of relatively long resonances would be, sometimes, broken due to frequency shift of a component. That would produce abrupt valleys on the curves. To solve this issue, a mechanism was developed in order to keep the magnitude components based on the existence of neighboring bins. The spectrogram of the results can be seen on Figure 3.7.

3.2.3 Hilbert Transform

The *Hilbert Transform* is a linear operator that results in the convolution of a real valued signal with the signal $-1/(\pi t)$. The process is a basic tool in Fourier analysis and can calculate the harmonic conjugate of a given function or Fourier series. Additionally, it can also be obtained on the time domain based on a moving average approach [24]. If this conjugate is added back to the original signal and processed using a LPF it is possible to extract the original signal's envelope.

The signal was segmented into 50% overlapping chunks (of 31.25ms) which underwent an

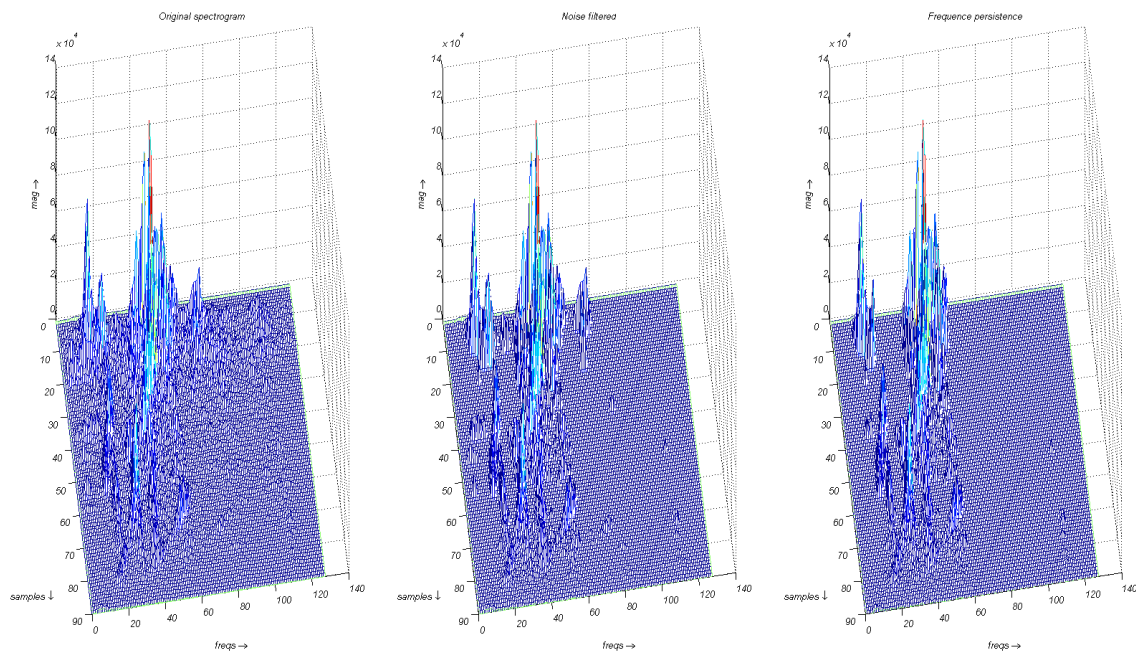


Figure 3.6: Frequency tracking spectrogram process, from left to right, original, noise filtered, persistence of resonances

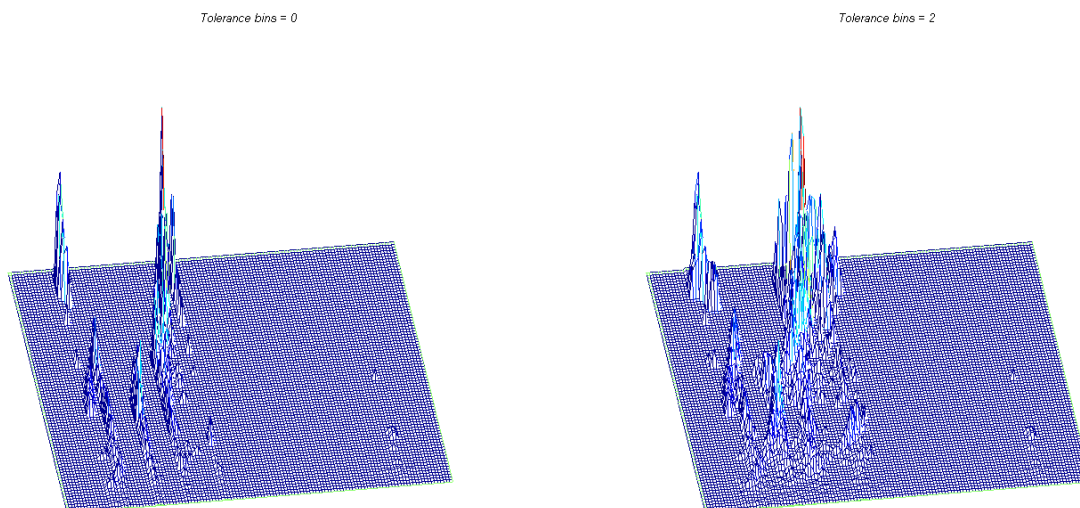


Figure 3.7: Frequency tracking spectrogram output, on the left using bin tolerance of 0, on the right using bin tolerance of 2

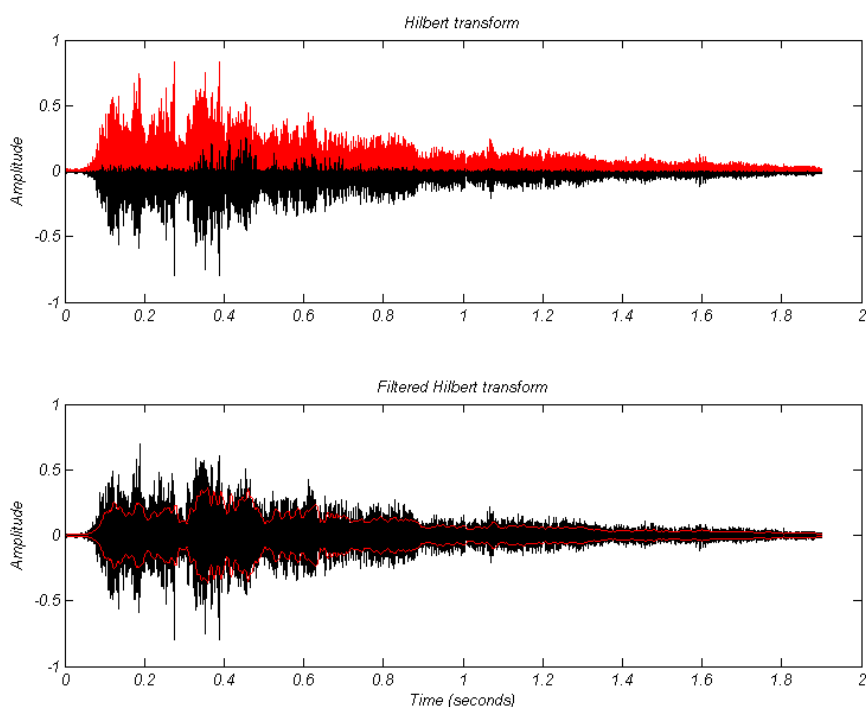


Figure 3.8: Original signal (in black) and Hilbert transform envelope (in red): Top image is an unprocessed result. Bottom image is the filtered result

FFT operation. The Hilbert Transformation was implemented using the Rice–Dugundji representation:

$$H_{\eta}(f) = \begin{cases} 0, & \text{for } f = 0. \\ -iF_{\eta}(f), & \text{for } f > 0. \\ iF_{\eta}(f), & \text{for } f < 0. \end{cases} \quad (3.5)$$

Where f corresponds to frequency, i is the imaginary unit, $F_{\eta}(f)$ is the Fourier series of the original audio signal, $H_{\eta}(f)$ is the harmonic conjugate, on the frequency domain. Finally, after the conjugate was calculated, the envelope was obtained.

3.2.4 Shannon Envelopes

The Shannon envelope functions, namely Entropy and Energy, come from the Information Theory field, employing notions of measure of uncertainty and expected value. These functions are quite popular for retrieving the envelope of AC signals in signal processing, particularly for enabling the segmentation/detection of cardiac sounds [25–28]. In some cases, the energy is normalized

and averaged [29].

$$\text{Shannon Energy: } E = -x^2(t) \cdot \log x^2(t) \quad (3.6a)$$

$$\text{Shannon Entropy: } E = -|x(t)| \cdot \log |x(t)| \quad (3.6b)$$

The equations 3.6a and 3.6b have slightly different characteristics. E is either the Shannon Energy or the Entropy and x is the input audio signal. As shown in Figure 3.9, Shannon Entropy focuses on the lower-mid intensities while the Energy on the mid-higher ones, respectively attenuating the counterparts. After the envelopes are obtained, they are either smoothed with a shifting mean or filtered.

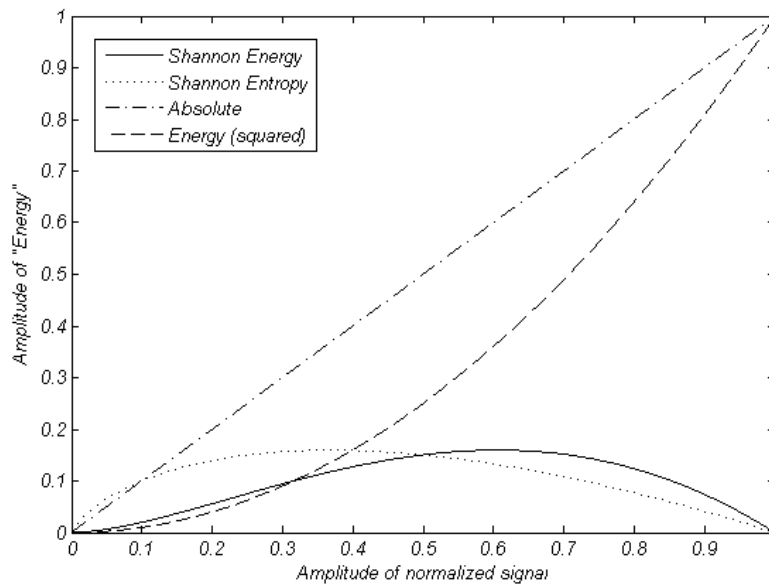


Figure 3.9: Comparison of different envelope methods. The solid line is the Shannon Energy that converts the original amplitude values and enhances the upper amplitudes. The dotted line is the Shannon Entropy that converts the amplitude values and enhances the lower amplitudes. The dashdot line is the absolute value ($|x|$) that has a linear amplitude response. The dashed line is the Energy (x^2) which is non-linear and boosts the dynamic range of the signal.

3.2.5 Linear Predictive Coding

Linear Predictive Coding (LPC) is a popular audio processing tool used to characterize a signal source in a compressed format. It is used for reducing speech information, subsequent transmission and voice reconstruction in a very accurate way, at the receiving end. There are many variants of LPC, namely LPC-10, CELP, MELP, RELP, among others. However, all of them are based

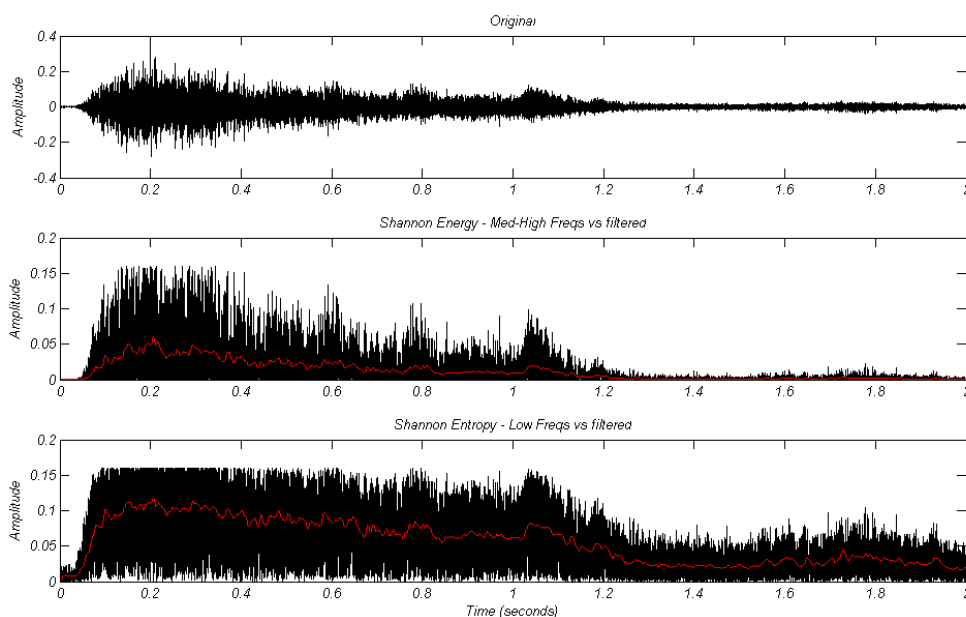


Figure 3.10: Comparison of Original signal and Shannon envelopes (black) and filtered results (red)

on a single model: an excitation signal and a filter. This speech/signal production model can be represented as a signal produced by a difference equation (3.7).

$$y(n) = \sum_{n=1}^p a_i y(n-i) \pm Gx(n) \quad (3.7)$$

Where $x(n)$ is the excitation signal, $y(n)$ the speech/output signal, G the "history parameter" and a_i the filter coefficients. The excitation signal can be either an impulse train, for voiced sound, or white noise, for unvoiced.

The respiration process implies two concepts similar to LPC: a vocal tract which can be modeled as a varying width tube [30]; and the exhalation itself, which is the power source of the expiration sound. This tube model is represented by the aforementioned filter and the power source can be modeled as white noise.

In order to retrieve the envelopes from the exhalation recordings, one can segment the input into 50% overlapping chunks of 31.25ms, and obtain the white noise variance, or power, from the LPC outputs. This succession of variance values should be proportional to the expiration power at the respective time [19], and thus, can be considered as a downsampled envelope of the original. The implementation exported the LPC envelopes with degrees 2, 4, 8, 16 and 32 to model increasing vocal complexity.

3.2.6 Post processing

Savitzky-Golay filter: The method popularized by Savitzky and Golay's convolution coefficient tables is based on the notion of weighted moving averages. The main drive of this method was the necessity to remove random noise from generic physics and chemistry related information, surpassing the contemporary methods in simplification and computation speed [31]. Once a convolution function is chosen, more precisely, after the degree of the a polynomial and the window size are chosen, the values of the window are convoluted with the polynomial coefficients, in order to calculate the window's center smoothed value.

3.3 Feature Extraction

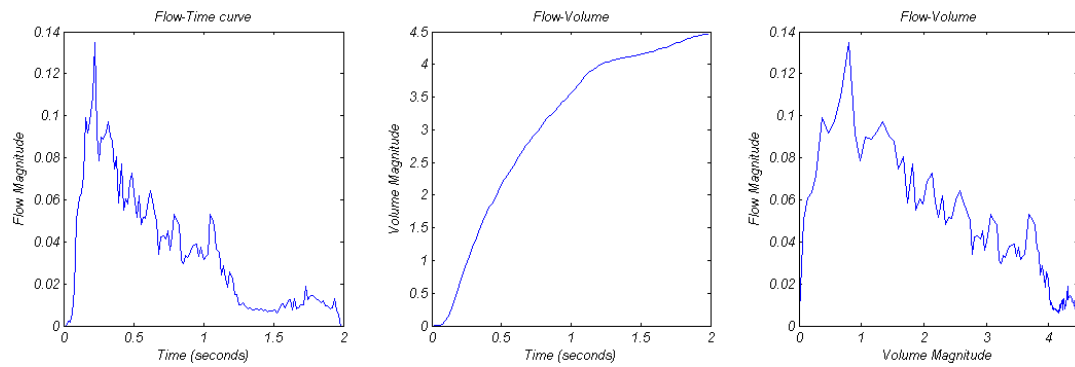
For each envelope the clinical parameters were extracted. This envelope represents the Flow-Time curve. However, some parameters can only be collected from the Volume-Time curve and, generally, a spirometer also presents the Flow-Volume curve. The Volume-Time curve can easily be generated by integrating the flow over time. The Flow-Volume curve is produced by remapping the Flow-Time points (abscissas) to the Volume-Time ordinates. As a consequence of constant rate sampling and a non-linear Volume-Time curve due to the FEM itself (specially at the beginning and end), the mapping presents uneven spacing between data points. This can be adjusted with simple linear or quadratic interpolation. Examples of these curves are shown in Figure 3.11.

The parameters are obtained as described:

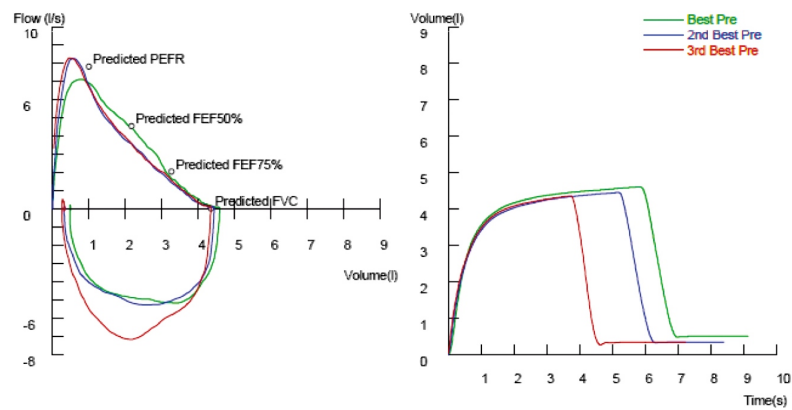
- The FVC corresponds to the total volume exhaled during the FEM, so it corresponds to the last value of the Volume-Time curve.
- The FEV₁ is the volume exhaled during the first second. It is equal to the value of the Volume-Time curve at the sampling frequency value sample.
- The FEV₁/FVC ratio is calculated as defined.
- The PEF corresponds to the maximum value of the Flow-Time curve.
- The parameters FEF_{25%}, FEF_{50%} and FEF_{75%} are gathered from the Volume-Time curve at the instants of 25%, 50% and 75% of the total volume, respectively.
- The FEF_{25%–75%} corresponds to half of the total volume (FVC) divided by the time difference between the 75% and 25% of the volume.
- Cust1 is calculated as $FEF_{25\%–75\%}/FEV_1 \times 100$.

As mentioned on Section 3.1.1, all the values extracted are not calibrated and thus, there is a need for a regression phase, to both reduce the number of features and adjust them to medical values range.

³Adapted from <http://www.associatedhealthsystems.com/images/pages/diagnostics/SpiroChart.jpg>



(a) Recorded data results: from left to right, Flow-Time, Volume-Time and Flow-Volume.



(b) Part of spirometer report³, three FEM maneuvers: from left to right, Flow-Volume and Volume-Time.

Figure 3.11: Clinical curves

3.4 Machine Learning

The dataset is constituted by a very unbalanced, small quantity of observations. There are significantly less cases of anomalous lung behavior (obstruction and restriction classes) than normal ones. Additionally, due to the feature space initially being considerably larger than the number of observations, it should also be considered the Hughes effect. This consists of the tendency of the classification error to increase as the feature space increases, for the same number of observations [32]. The dataset will be discussed in Section 4.2. This conjugation of factors could easily enable a situation where the classifier would overfit. Therefore, it was decided to employ a regression-first-classification-later approach. A discussion concerning the choosing of the implemented methods is presented on Section A.2.

3.4.1 Regression Algorithms

Regression Trees: A tree is a structure that consists of a root node and several others attached to it in some kind of hierarchical fashion. The decision tree's nodes represent decisions, separating data according to the node's function. The terminal leaf nodes are the data labels. This kind of structure is transparent to the user since it progressively sorts the data based on some feature until it is completely defined.

The tree, or in fact, the node's functions, are generated based on the training data. There is significant effort to find the best and fastest ways to divide the training data. The termination of the dividing algorithm generally isolates many data entries to their own leaf. This causes problems concerning the overfitting of the method. Many variations were proposed to reduce this problem such as terminating the algorithm before the tree grows to full depth or pruning some of the outlier branches.

Specifically, the terminal nodes of Regression Trees consist on the attribution of the average value of the training data that falls into it. Each new data entry undergoes the decision process, traversing through the tree until the leaf node is reached and the respective numeric value is assigned.

3.4.2 Classification Algorithms

Classification Trees: are quite similar to Regression Trees, except for the fact that the leaf nodes' value does not correspond to a numeric value but to the most frequent classification of the training set that arrived to that node.

Support Vector Machines (SVMs): SVMs [33] are focused on separating labeled data using hyperplanes. Their aim is to maximize a "margin" between the boundary of different classifications

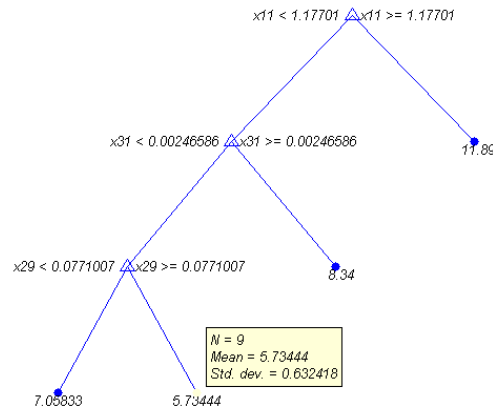


Figure 3.12: Regression Tree example, regressing PEF

and the instances of either side. Minimizing the Lagrangian in 3.8 with respect to the weight vector w and bias b would get the best hyperplane cut [3].

$$L_P \equiv \frac{1}{2} \|w\|^2 - \sum_{i=1}^N \alpha_i y_i (x_i \cdot w - b) + \sum_{i=1}^N \alpha_i \quad (3.8)$$

where x is observation matrix, y is the observation label vector, α is the observations' Lagrangian multiplier vector. The SVMs' training algorithms necessarily reach the optimal separation, avoiding local minimums. Though, if each dimension is to be considered equally important, i.e. have the same weight, and enable a more generic solution, the dataset should be normalized before being fed to the training model.

When the data is not linearly separable there are some non mutually exclusive methods to reach a solution. Usually, even if some data points are misclassified or fall within the other classes' hyper-dimensional region a classification model can still be devised by tolerating the misclassification of some data points. These are called *soft margin* SVMs. By introducing positive slack variables (ξ_i) on the constraints it is possible to accept some misclassifications of the training set. The Lagrangian turns into:

$$L_P \equiv \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \xi_i - \sum_{i=1}^N \alpha_i \{y_i (x_i \cdot w - b) - 1 + \xi_i\} + \sum_{i=1}^N \mu_i \xi_i \quad (3.9)$$

where μ_i are the Lagrange multipliers that keep ξ_i positive.

The second method solves the problem of simple non-linear separation as a transformation of the dataset, named kernels. For instance, if a dataset had a radial distribution, like the one presented on Figure 3.13, it could be linearly separated provided the radial basis function kernel would be applied before the SVM was trained.

Finally, one can also improve the SVM classifier performance for not linearly separable data by adding a dimension to the dataset [3].

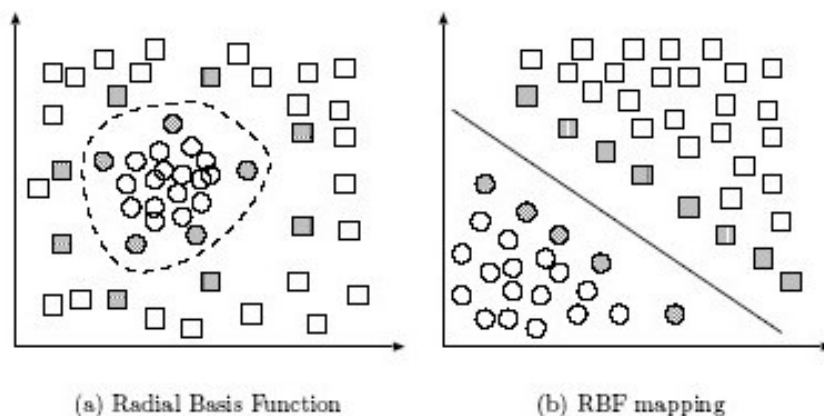


Figure 3.13: Separable classification with Radial Basis kernel functions in different spaces. Left: original space. Right: feature space⁴

s

Although, SVMs are strictly binary classifiers there are also strategies to cope with multiple classes at the same time by dividing the problem into several suboptimal ones. Some are, for example, One against One (also referred as All against All) and One against All. The first strategy consists of developing one SVM classifier for each pair of classes while the later consists of developing N SVM classifiers where each distinguishes between one class and the remaining others. In either case, the class with most votes is deemed the output [34].

Naïve Bayes: is a predicting model that, given an n dimensional distribution θ , can compute a class' probability given new data, $P_{\theta}(C|D)$. The model's decision should be the one with highest conditional probability. According to Bayes rule, the conditional probability can be rewritten as in equation 3.10a. However, since the maximization will be made over C , $P(D)$ can be ignored. Also, it is assumed that the data features are conditionally independent given the classification, enabling the conditional probability to be rewritten as equation 3.10b. Both likelihood, $P(D_i|C)$, and prior, $P(C)$, can be estimated from the training set and so, the model output is given by equation 3.10c

$$P(C|D) = \frac{P(D|C)P(C)}{P(D)} \quad (3.10a)$$

$$P(D_1, \dots, D_n|C) = P(D_1|C) \dots P(D_n|C) \quad (3.10b)$$

$$\hat{C} = \arg \max_C (P_{\hat{\theta}}(D|C)P_{\hat{\theta}}(C)) = \arg \max_C \left(P_{\hat{\theta}}(C) \prod_{i=1}^n P_{\hat{\theta}}(D_i|C) \right) \quad (3.10c)$$

⁴Image taken from <http://www.dtrek.com/svm.htm>

3.4.3 Ensemble Algorithms

Bootstrap Aggregation: A known technique of variance reduction that diminishes overfitting problems is Bootstrap Aggregation (Bagging). By using n random data subsets, with replacement, to generate learner models and enabling a voting/averaging process of new data, it can enhance local adaptivity of the boundary [35]. It works well with classifiers that are unstable like decision trees.

Random Forests: This method is similar to Bootstrap Aggregation, however, it differs on the number of observation features/dimensions that are considered at each split of a decision tree. An initial random training subset is used to grow each tree and, at each node, only a random subspace of features of the training set is used for splitting. Eventually, each tree is fully grown (not pruned) and either the majority vote (classification) or the average (regression) of all trees decision values is deemed the method output for a given input. This method is relatively robust to outliers and noise, and is faster than either bagging or boosting. In terms of accuracy, it also outperforms bagging and rivals with boosting [36]. Random Forest number of deciding features usually are one third of the total number, for regression, and the square root of the total number, for classification.

Boosted Methods: The boosting process is an online learning method that enables weak learning models, such as decision trees, to become high performance models. The concept of boosting comes from the idea of focusing the learning process on the hard-to-learn portion of a distribution [37]. Like any other online learning method. The learner is first presented with a training set with uniform importance, tries to learn the model and how to classify observations. The weak learner model is stored, each wrong observation weight is increased and a new weak learner model is trained the same way, using the new distribution weights. After a fixed amount of iterations, a uniform voting process, using all the stored learners is used to decide the test data set.

Adaptive Boosting (AdaBoost): AdaBoost is an improvement of generic boosting methods whose voting process is weighted [38]. Each weak learner's vote weight $\alpha_i = \log\{(1 - e_b)/e_i\}$, where e_b is the iteration's weighted missclassification error given by equation 3.11:

$$e_b = \frac{\sum_{i=1}^n \omega_i 1 \{y_i \neq \hat{f}_{w,b}(x_i)\}}{\sum_{i=1}^n \omega_i} \quad (3.11)$$

where $\hat{f}_{w,b}$ is the b th weak learner from the ensemble class estimation, n is the number of training observations, x_i and y_i are the i th training data and classification, respectively, and ω_i is the i th observation weight. The curly braces indicate the expression is logical. Either w_i is multiplied by 0 or by 1.

There are two different implementations: AdaBoost.M1 is for binary classification and AdaBoost.M2 is for multi-class problems, though their details will not be approached in this document.

Chapter 4

Experimental Evaluations

This chapter describes the experiments conducted to evaluate how different groups of signal processing, regression and classification methods perform using the available dataset. The chapter is divided in experimental setup and the description and results of those experiments. First, the tools used are presented and the recording protocol explained. The data acquisition processes is described along with how the dataset is composed and how it was obtained. The experiments initially devised are presented, followed by those that were afterward deemed interesting in light of the first results. Finally, the architecture of the final prototype is proposed.

4.1 Experimental Setup

4.1.1 Technologies and tools

The methods and architecture were implemented in *Matlab*, as well as the simulations. The SVM models were implemented using LIBSVM.

The "ground truth" clinical values were obtained using the spirometer devices presented on Table 4.1. The recordings were made on a Samsung GT-I9000, using a custom made app designed by the student and developed by Bernardo Pinho.

4.1.2 Recording process/protocol

Firstly, each patient performs a few forced expiratory maneuvers into the clinical spirometer. This step was repeated until the ATS repeatability standards were achieved. From the best maneuver, the clinical parameters, mentioned on subsection 2.1.2 were obtained. Since most of the recording process was included in screening events, other tests were conducted, namely, Nasal Peak Inspiratory Flow (PIF) measurement, carbon monoxide (CO) quantity measurement and nitric oxide (NO) exhaled fraction measurement.

The patient was then explained how to perform the FEM with the smartphone, stretching his/her arm at mouth level, holding the device, facing the screen and not the microphone. It was



Figure 4.1: Screenshots from the enhanced recording app

emphasized to exhale directly to the screen, avoiding holding the phone either tilted or obstructing the microphone. A few recordings were made in order to reduce variability.

Afterwards, the patient was often asked to do a post-bronchodilator (BD) test, repeating the previous steps after a few minutes of taking the drug. Due to time constraints of the events or excessive effort/strain caused by the numerous FEMs, not all patient's were asked to record a post-BD maneuver.

In all maneuvers, using a spirometer or otherwise, the patient's were asked to wear a plastic nose-clip to prevent exhaling through the nose.

The recordings were made with the following settings: AMR file encoding, on a 3GP package, with a 16kHz sampling rate. All the recordings were later converted to Wave format at a sample rate of 16kHz and bit depth of 16 using the *ffmpeg* codec¹.

4.2 Dataset

The dataset is composed by 101 recordings from 61 patients performing the forced expiration maneuver, without any mouthpiece, at an arm's length. Additionally, each recording is accompanied by the patient's anthropometric parameters (age, height, weight, gender and race), the clinical parameters, and classification (classtype) and degree label (classdeg) of the patient's lung function provided by the recording physician.

The classification types are normal, obstruction, restriction and mixed, and the degrees are mild, moderate, severe and other. There are also annotations concerning the recording environment, who performed the recordings, an open text field (for indicating that a bronchodilator was used or there was coughing, for example) and a classification of the recording/maneuver as good,

¹www.ffmpeg.org/

bad cooperation or error (no maneuver performed). This last field provides a quick insight of the recording's content which is helpful since all recordings were saved for later analysis.

The recorded patients were either part of studies of Control and Burden of Asthma and Rhinitis (ICAR), patients from CUF Porto Institute (ICP) or from CUF Porto Hospital (HCP). The recordings were made on the dates and places shown on Table 4.1.

Table 4.1: Recordings' Distribution and ground truth device

Number of recordings	Date	Location	Spirometer
14	3-04-2014 to 5-04-2014	Aveiro - ICAR	Spirodoc - MIR
17	10-04-2014	Coimbra - ICAR	Spirodoc - MIR
18	22-04-2014 to 20-05-2014	ICP	Jaeger IOS – Carefusion
2	15-04-2014	HCP	Jaeger IOS – Carefusion
10	17-05-2014	Vila Real - ICAR	Spirodoc - MIR
9	24-05-2014	Viseu - ICAR	Spirodoc - MIR
15	3-06-2014	Loulé - ICAR	Spirodoc - MIR
16	4-06-2014 to 5-06-2014	Portimão - ICAR	Spirodoc - MIR

The patients ranged from 18 to 82 years old. Since children's lungs have higher elasticity and, therefore, very different characteristics than adults and teenagers, they were not considered for the study. The patient classification distribution is shown on Table 4.2.

Table 4.2: Recordings' clinical classification distribution

Class	Total	Mild	Moderate	Severe
Normal	71	-	-	-
Obstruction	23	21	2	0
Restriction	2	0	0	2
Mixed	5	2	3	0

4.3 Regression Experiments

The algorithms used for the subsequent experiments were based on a backward selection approach. Initially, all the signal processing methods were used and the clinical parameters' sets were obtained by successively removing some methods out of the initial set. For each set, a 5-fold cross validation set was made to verify the models' expected accuracy and to obtain the average error and standard deviation. All presented error values are shown in their absolute value. This also applies to boxplots and other figures. Some tables present method sets labeled using a negative, for instance, "Minus p_{lips} " or "No Energy". This tries to clarify that those functions are the only ones removed from the current envelope set, and that the remaining others are still combined.

4.3.1 Preliminary Regression Results

First, the stability of the regression ensemble methods was verified and the best number for weak learners chosen. Table 4.3 shows the results of the test. Both Bootstrap Aggregation (Bagging) and Random Forests were tested using 5-fold cross validation (out-of-bag) prediction.

Table 4.3: Regressed Values of FEV_1 , using Bagging and Random Forests, using the original signal processing architecture

Methods	Measurement	Number of Trees					Avg.	Std.
		80	110	140	170	200		
Bagging	Avg. Error %	22.28	21.41	21.25	20.78	20.82	21.31	0.61
	Error Std. %	21.08	20.42	21.51	20.05	20.53	20.71	0.57
	K Fold Loss	2.99	2.86	2.71	2.56	2.64	2.75	0.17
Random Forest (1/3)	Avg. Error %	21.72	21.84	22.63	20.60	21.04	21.56	0.78
	Error Std. %	21.03	22.03	22.65	20.39	20.48	21.32	0.99
	K Fold Loss	2.73	2.84	3.03	2.58	2.62	2.76	0.18

The preliminary results show that bagging and random forests are quite similar. Even though the results are very stable, and nearly identical, the best number of decision trees for each ensemble seems to be around 180. Henceforth, 180 trees are used for all regression models. Additionally, both methods consistently produced between 5 to 7 outliers, with regressed values reaching the double of the intended targets (from the spirometer).

4.3.2 FEV_1/FVC Regression and Calculation

Since the regression methods try to model as close as possible the conversion from data envelope parameters to the target values, it stands to reason that the FEV_1/FVC parameter is closest when it is directly regressed than when it is calculated based on the regressed values of FEV_1 and FVC. Also, as a consequence of regression, both values of FEV_1 and FVC are prone to error and the error of each one can be of different sign, diverting considerably the ratio.

In order to test that assumption, a complete parameter set, containing all methods, as initially described on Figure 3.1, was generated. Since the regression should be fairly the same for this number of envelope methods with either bagging and random forests, it was decided to compare the index's values using random forests, that enable a shorter learning time. The learning and prediction procedure was the same as in the previous experiment, 5 fold cross validation, with 180 trees in each forest. Figure 4.2 presents the results of the test. It clearly supports the original thesis. The directly regressed values of the index were used henceforth.

In order to try to simplify the tables and diagrams, the parameter the FEV_1/FVC index is referred as *Tiff*, hereafter. It is not exactly the Tiffeneau's index, in a medical sense, since the Tiffeneau's index is defined as FEV_1/VC , as described on Section 2.1.1.

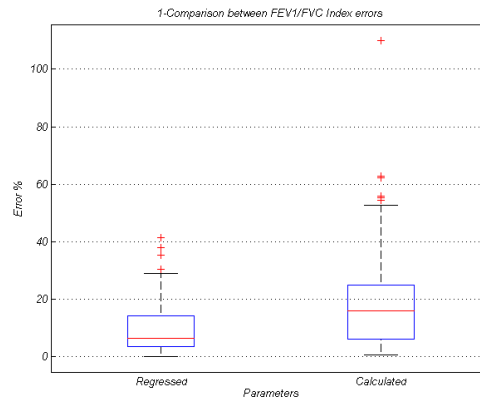


Figure 4.2: FEV₁/FVC index error comparison, to the left, directly regressed values, and to the right, calculated based on regressed values of FEV₁ and FVC, independently

4.3.3 Comparison of Conversion vocal models

As discussed on Section 3.2.1, the Inverse Radiation Model (IRM) can be reduced to a gain. This raises the question if the respective waveforms should be used for envelope generation, since the envelopes should not change their relative values, i.e., the curve shapes should be the same. This means that including the results from this model would introduce redundant information on the learning process and, most likely, would increase the learner's overfitting.

This experiment involves comparing the regression results from three envelope sets. The first set (All) includes all the methods initially considered for the systems, as described in Section 3.1. On the second set (Minus p_{lips}), the envelopes generated from the p_{lips} waveform (IRM's output) are removed. On the third set (Minus p_{lips} gain), the influence of the IRM is removed from the Pressure to Flow Conversion Model (PFM). Specifically, the PFM's input becomes directly the recording's audio.

Table 4.4: Conversion Models experiment - Regressed Values Error - Bagging

Models		Regressed Parameters error								
		PEF	FEV ₁	FVC	Tiff	MMEF	CustI	MEF25	MEF50	MEF75
Avg.%	All	20.91	30.40	20.05	9.99	42.16	25.10	71.23	41.81	37.46
	Minus p_{lips}	21.93	29.11	19.55	9.76	41.19	24.86	68.92	40.84	33.36
	Minus p_{lips} gain	21.41	29.58	21.14	9.84	42.96	25.02	70.79	40.91	35.33
Std.%	All	20.04	27.78	18.10	8.78	53.53	27.80	100.16	46.87	47.06
	Minus p_{lips}	22.29	25.58	19.10	8.97	53.04	27.17	96.03	47.04	44.40
	Minus p_{lips} gain	22.27	26.77	20.14	8.56	56.69	28.05	105.07	46.12	41.25

where Avg. corresponds to the percent average of regression error, Std. is the percent standard deviation of regression error and Tiff corresponds to the FEV₁/FVC index.

Both Tables 4.4 and 4.5 and Figure 4.3 present the regression results of the three method sets across the 9 clinical parameters.

Table 4.5: Conversion Models experiment - Regressed Values Error - Random Forest

Models		Regressed Parameters error								
		PEF	FEV ₁	FVC	Tiff	MMEF	Cust1	MEF25	MEF50	MEF75
Avg.%	All	22.70	30.92	20.99	9.74	41.96	25.30	69.28	42.04	34.06
	Minus p_{lips}	22.03	31.38	20.44	10.09	42.08	24.63	64.96	38.87	35.26
	Minus p_{lips} gain	20.84	29.42	21.41	9.86	40.40	24.90	71.14	39.32	35.01
Std.%	All	21.01	29.02	19.03	8.46	54.32	28.28	92.81	44.41	41.54
	Minus p_{lips}	21.59	28.67	19.47	9.30	52.35	26.74	86.91	44.57	45.00
	Minus p_{lips} gain	19.99	27.97	20.34	8.79	50.68	27.09	100.55	42.54	44.18

One relevant analysis that can be conducted over these results comprise each parameter's regression variability and their relevance to medical diagnosis. As Figure 4.3 shows, frequently used lung function indicators as FEV₁/FVC (shown as Tiff), FVC and PEF rank among the ones with the least variability, denoted with lower 75 percentile. Both MMEF and MEF_x parameters also followed the predictions of high variability since those values are rarely medically used for those same reasons. On the other hand, FEV₁ was not expected to have such high variability than the aforementioned parameters, since it is widely used by physicians. The custom parameter, Cust1, managed to present some cautiously promising results. Though, it ought to have greater meaning and impact on the classification, since no other study has established the characteristic ranges of the parameter, to the best of our knowledge. Either way, the relatively small error of FEV₁/FVC (~10%) presents a good starting point for later classification.

The results show that, for both Bagging and Random Forests, either method set is similar to the others, in terms of regression error. In order to reduce by one third the number of curves for which to calculate the clinical parameters, the conversion model $p \rightarrow p_{lips}$ was removed. Additionally, that model's gain, also does not seem to produce any changes. Therefore, the gain will be ignored.

To support the removal of certain envelopes from the set, another test was devised. 5 fold ensembles of 180 trees were trained. Each variable/generating envelope split was counted over the all the trees and all the 9 clinical parameters. A gross approximation of the importance of an envelope method can be obtained by comparing the total number of splits. Figure 4.4 shows the

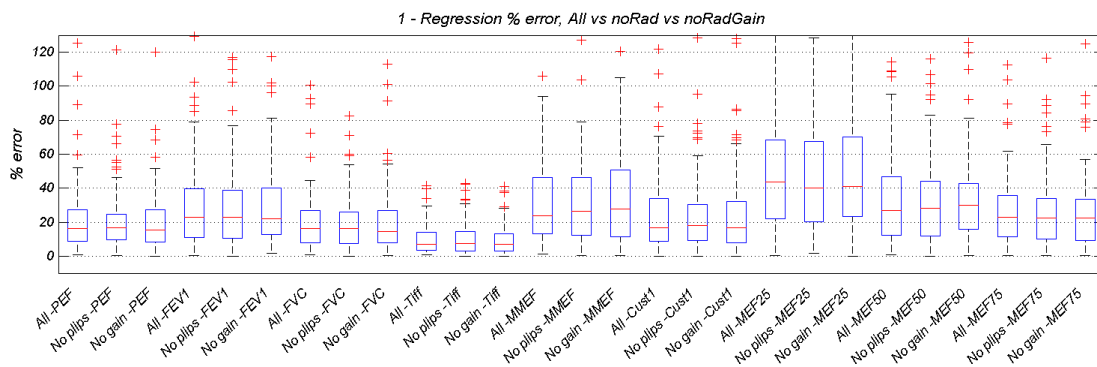


Figure 4.3: Comparison of Bagging Regression Error Distribution along the clinical parameters and sets of methods (enlargement)

results for the complete set of envelopes, as originally described.

The labels *shanEnerg* and *shanEntrp* refer to the envelopes retrieved by Shannon Energy and Entropy, respectively, *Hilbert* to Hilbert Transform, *LPC 2* through *32* refer to the Linear Predictive Coding, with the respective order and *mRes* refers to Mean of Resonances. Additionally, *LP* indicates that the envelope underwent a Low Pass filter, *sGolay* means it underwent a Savitzky-Golay filter and *poly* that the envelope was fitted using a polynomial, all of which are described on Section 3.1.1.

It is straightforward to see that the methods with less split count generally are those of *pLips*, thus corroborating the previous findings. In fact, the lowest 20% of the summed splits over all 9 parameters, are all based on *pLips*, except for one from *uLips* (LPC 16-polynomial fitting) and two from *p* (LPC 4 and Hilbert - polynomial fitting).

4.3.4 Filtering options

The output of the envelope generating methods is considerably noisy and quite different from a spirometer's curve, in terms of smoothness, thus a filtering algorithm is needed. Two options were devised as described on Section 3.1.1, either using Butterworth Low Pass filters or Moving Average. Both were tested, in mutual exclusivity, on the same set of methods (same as before, without *pLips* model or gain) and the error results are shown on Figure 4.5 and Tables 4.6 and 4.7.

Table 4.6: Filtering options experiment - Regressed Values Error - Bagging

Models		Regressed Parameters error								
		PEF	FEV ₁	FVC	Tiff	MMEF	Cust1	MEF25	MEF50	MEF75
Avg.%	LPF	20.22	29.85	22.51	9.94	39.76	26.49	70.07	40.09	35.09
	MA	21.92	30.07	21.99	10.04	43.44	25.68	70.87	38.80	33.72
Std.%	LPF	22.08	25.70	21.40	8.90	47.93	27.49	101.52	43.36	46.22
	MA	21.28	25.99	21.60	9.06	56.29	27.13	88.86	45.76	44.11

Table 4.7: Filtering options experiment - Regressed Values Error - Random Forest

Models		Regressed Parameters error								
		PEF	FEV ₁	FVC	Tiff	MMEF	Cust1	MEF25	MEF50	MEF75
Avg.%	LPF	22.19	30.07	22.16	9.65	42.49	24.30	69.25	39.74	36.14
	MA	21.75	29.61	21.73	10.12	42.17	25.80	66.78	42.09	36.33
Std.%	LPF	22.49	27.01	21.25	8.80	55.70	26.64	91.18	43.69	40.69
	MA	20.42	25.92	20.42	9.32	51.39	27.43	87.01	47.07	47.79

The results are quite similar, leaving no obvious criteria for deciding which set of filtering methods to employ. The Low Pass Filter approach was chosen since the Moving Average (MA) recurrently uses division, becoming more computationally intensive than LPF, which uses only multiplication by pre-calculated coefficients. Additionally, the Savitsky-Golay filter is based on a weighed moving average approach. Consequently, it seems counter-productive to continue using the Moving Average since it will be afterwards smoothed in a similar fashion.

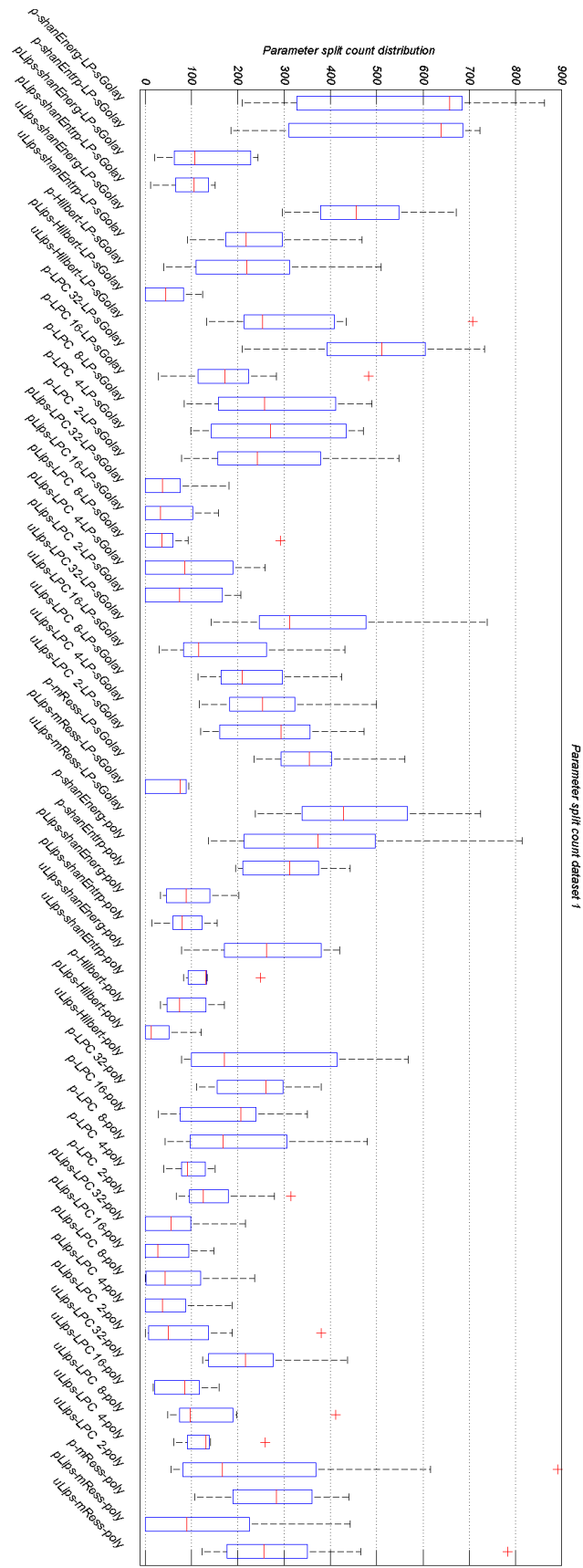


Figure 4.4: Split count distribution over the 9 parameters, for an envelope set containing all methods

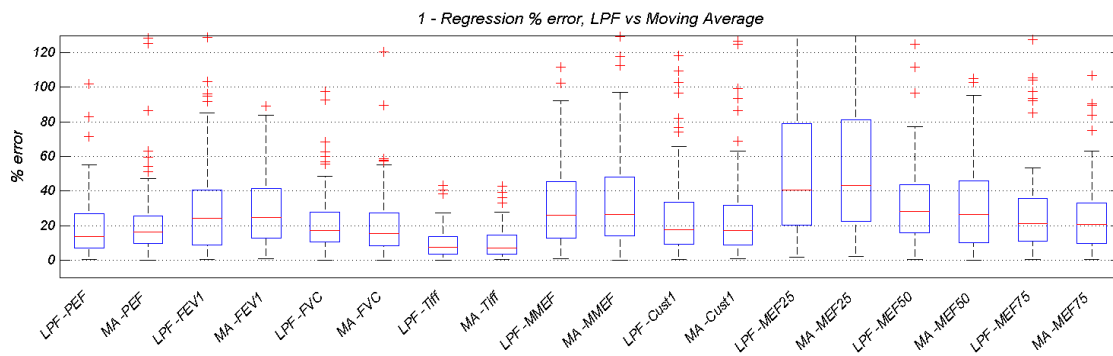


Figure 4.5: Comparison of Low Pass Filter and Moving Average, Bagging Regression results. Error Distribution, throughout the clinical parameters, using the same set of methods (enlargement)

4.3.5 Hilbert Transform and Shannon Envelopes

The SpiroSmart project [19] uses Hilbert Transform, as a general envelope calculation method, Mean of Resonances, which focuses on the magnitude of the frequency components characteristic to each patients vocal tract, and Linear Predictive Coding, which models the patients vocal tract with filters and, in this case, only the sound source's power is retained. All of these methods try to develop envelopes using somewhat different concepts.

For this dissertation, the Shannon Envelopes are proposed as a general envelope calculation method, similarly to the Hilbert Transform. However, as each of the Shannon envelopes focuses on different intensities, they should be compared as a pair. Comparing these general envelope generators was considered relevant since they serve the same purpose and the best solution should employ the least number of methods as possible, to improve energy efficiency and make the system computationally lighter. Figure 4.6 and Tables 4.8 and 4.9 show the results of both methods used together, removing only Hilbert Transforms and removing only Shannon envelopes.

Table 4.8: Hilbert Transform and Shannon Envelopes experiment - Regressed Values Error - Bagging

Models		Regressed Parameters error								
		PEF	FEV ₁	FVC	Tiff	MMEF	Cust1	MEF25	MEF50	MEF75
Avg.%	All	20.19	31.67	21.33	9.74	40.87	25.74	68.99	43.20	35.28
	No Hilbert	21.61	29.66	21.45	10.03	40.64	23.90	71.57	38.93	32.71
	No Shannon	21.99	31.31	22.61	9.45	39.56	25.50	64.85	41.78	33.28
Std.%	All	20.91	28.52	21.75	8.89	52.79	26.23	95.51	51.46	48.69
	No Hilbert	22.70	27.07	20.36	8.95	52.94	26.79	97.65	41.84	40.89
	No Shannon	21.82	26.85	21.02	8.70	50.88	28.03	86.11	47.46	41.91

The results presented on the tables and on Figure 4.6 are similar across the three sets of methods and all parameters. As with the comparison between Low Pass filtering and Moving Average, there is no clear method set that confers a particular advantage over the others, including using both methods (*All* method set).

Since one of the objectives of the project is to keep a reduced number of methods, it was considered that the Shannon envelopes should be kept instead of the Hilbert Transform. Although,

Table 4.9: Hilbert Transform and Shannon Envelopes experiment - Regressed Values Error - Random Forest

Models		Regressed Parameters error								
		PEF	FEV ₁	FVC	Tiff	MMEF	Cust1	MEF25	MEF50	MEF75
Avg.%	All	21.92	31.16	22.76	9.64	42.62	25.00	69.84	41.63	39.34
	No Hilbert	21.39	30.83	21.27	10.11	41.64	24.15	69.52	39.33	36.06
	No Shannon	21.81	30.50	22.48	9.92	42.08	24.25	64.74	42.26	33.76
Std.%	All	21.24	28.54	22.41	8.82	55.61	27.06	89.86	52.34	61.11
	No Hilbert	21.90	27.55	18.55	8.94	51.01	26.36	89.75	42.60	43.63
	No Shannon	20.36	26.12	20.85	9.11	52.15	26.71	84.65	46.47	40.99

this means having to double the envelopes to filter, smooth and extract for parameters than just using the single Hilbert Transform, it is reasonable to say that it would still be the best option in terms of speed and computational complexity. Hilbert Transform implies taking the FFT and IFFT of the signal while the Shannon envelopes just square and take the envelope's absolute values and compute the base 10 logarithm. For a current smartphone with a dedicated floating point co-processor (which is fairly common) removing the Hilbert Transform envelopes would be the best option.

Therefore, for the remaining experiments the Hilbert Transform envelopes were removed.

4.3.6 Shannon Entropy and Energy

As mentioned on Section 3.2.4, Shannon Entropy enhances the lower intensities while Shannon Energy emphasizes the mid-higher intensities. The proposed experiment aims to verify if either has better results than the other or if both should keep being combined together to provide a more accurate regression.

Tables 4.10 and 4.11 and Figure 4.7 show the results of regression on sets with both methods or either without Shannon Entropy or Energy. As shown, removing either envelope method does not increase significantly the error's average or standard deviation. In fact, all three envelope sets present very similar results, across all parameters. This does not make clear which of envelope functions should be removed.

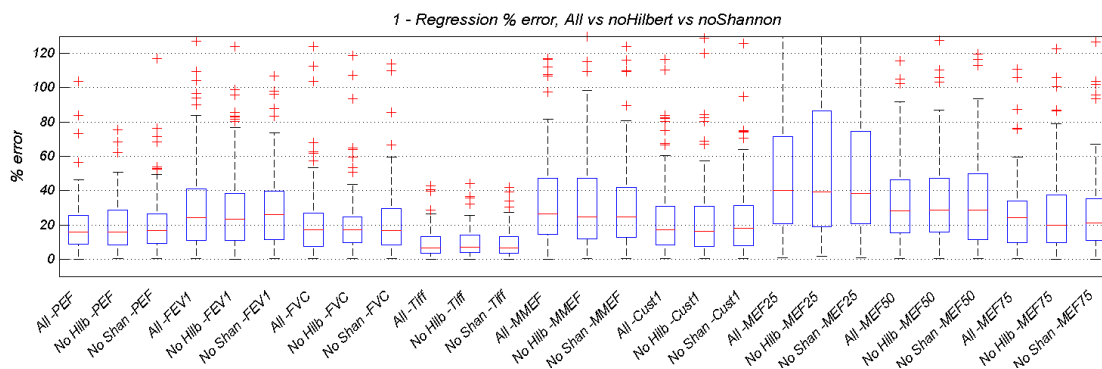


Figure 4.6: Comparison of Shannon envelopes and Hilbert Transform envelopes' influences on Bagging Regression results. Error Distribution, throughout the clinical parameters (enlargement)

Table 4.10: Shannon Entropy and Energy experiment - Regressed Values Error - Bagging

Models		Regressed Parameters error								
		PEF	FEV ₁	FVC	Tiff	MMEF	Cust1	MEF25	MEF50	MEF75
Avg.%	All	23.03	31.55	22.61	10.18	41.24	25.53	71.24	40.29	33.79
	No Entropy	20.93	29.97	23.54	9.75	44.92	25.04	72.74	39.08	34.20
	No Energy	21.53	31.36	21.79	9.78	40.96	24.28	69.62	37.96	36.10
Std.%	All	21.38	29.90	20.77	8.87	51.20	26.58	96.71	43.74	44.35
	No Entropy	20.10	25.90	24.12	8.71	58.57	26.81	85.23	41.31	47.02
	No Energy	23.37	27.15	21.29	8.89	57.04	25.51	91.32	44.83	48.31

Table 4.11: Shannon Entropy and Energy experiment - Regressed Values Error - Random Forest

Models		Regressed Parameters error								
		PEF	FEV ₁	FVC	Tiff	MMEF	Cust1	MEF25	MEF50	MEF75
Avg.%	All	22.96	29.67	21.88	9.72	42.27	25.41	69.76	39.01	36.44
	No Entropy	21.48	30.13	23.04	9.92	40.29	25.36	71.56	40.18	34.23
	No Energy	21.73	30.12	23.34	10.17	42.11	24.08	69.67	40.71	35.33
Std.%	All	25.39	26.39	22.41	9.01	54.91	27.73	90.19	44.75	56.35
	No Entropy	21.06	27.11	21.04	9.01	50.87	26.61	93.51	45.29	42.82
	No Energy	22.13	27.79	24.37	8.84	55.42	25.92	85.26	45.06	43.31

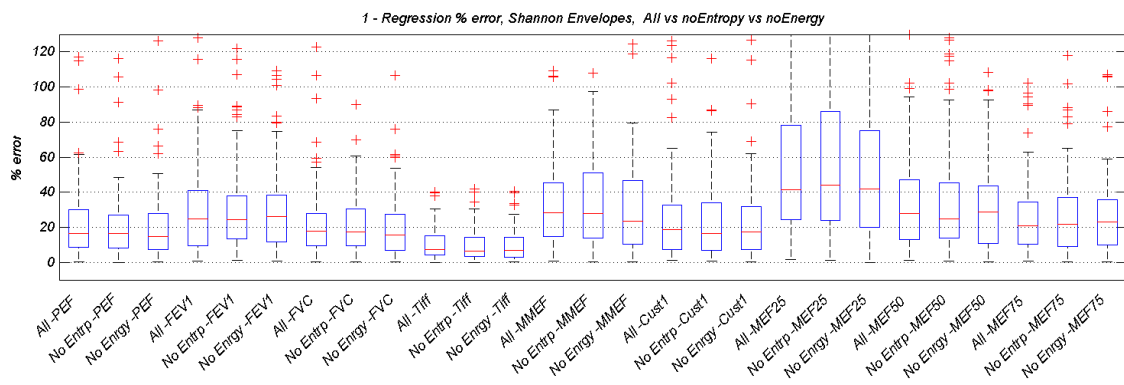


Figure 4.7: Comparison of Shannon Energy and Entropy envelopes' influences on Bagging Regression results. Error Distribution, throughout the clinical parameters (enlargement)

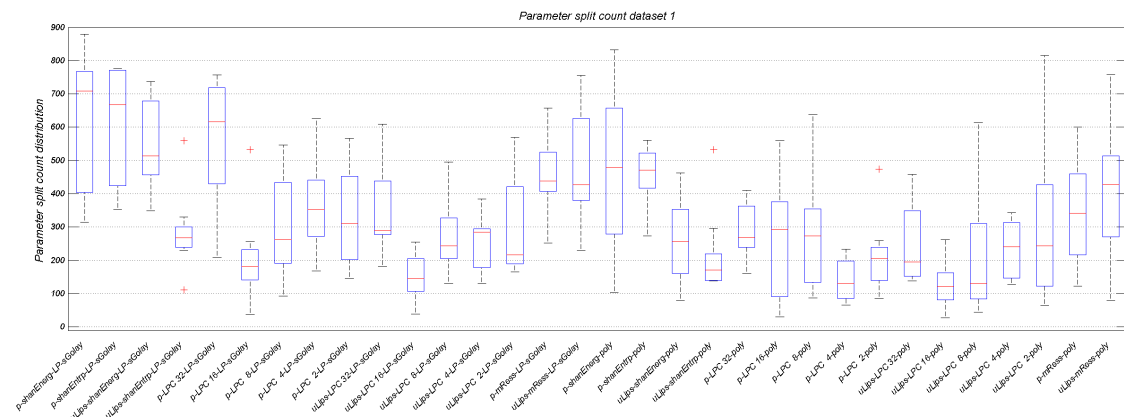


Figure 4.8: Split count distribution over the 9 parameters, for a set containing both of Shannon envelope methods

Figure 4.8 presents the relative gross importance of each method on the parameters' regression process. At this stage, the total number of methods is considerably less than the original set. The figure shows that among the two Shannon options, the one with the most frequently high split count is Shannon Energy. Let's consider each pair of Shannon envelopes within each processing option. In the *p-SGolay* process, both Energy and Entropy are quite similar and have a split count distribution that is fairly the same, however, when comparing the *u_{lips}-sGolay* or *u_{lips}-poly* processes, Energy seems to have a more consistent distribution than Entropy.

Considering both the similarity between the error results and the slight one-sided overuse on the split count distribution, the Shannon Entropy envelopes are henceforth removed from the envelope set.

4.4 Additional Regression Experiments

The first set of experiments enabled to trim a significant amount of envelopes and, consequently, remove 9 times more parameters from regression analysis and learning. The first objective of this project is to accurately retrieve the lung function indicators from the available envelopes and not the best, medical grade, exhalation envelope. Therefore, several of the remaining methods still seem redundant for the goal task. Such methods include the five grades of LPC envelopes and the Savitzky-Golay smoothing.

In this section, further experiments are presented as well as their results. The first experiment consists of removing some of the LPC envelopes, the second concerns trading smoothing after filtering for just filtering. The last experiment seeks to perceive the importance of the polynomial approximation for the regression.

4.4.1 LPC fundamental vocal complexity

Figure 4.8, can still provide an insight on which methods are important to correctly regress the parameters. Among the methods with the lowest percentages of total split count are some LPC envelopes of specific orders 2, 4, 8 and 16. To be precise, all of LPC 16 envelopes seem to have relatively low split count and only some of the envelopes of orders 2, 4 and 8 generate a high split count. In general, the envelopes of order 32 are more likely to generate splits. For simplicity's sake, a test was made, "downsampling" the LPC's orders, comparing LPC 2, 8 and 32, named LPC set A, and keeping only those that model the most and least complexity of the vocal tract, namely orders 2 and 32, LPC set B.

Tables 4.12 and 4.13 and Figure 4.9 show the results of removing these LPC envelopes from the current group, set A and B. The results present no significant change across all parameters, when removing first, orders 4 and 16, and second, order 8 LPC envelopes from the method set. This goes to show that these additional envelopes were indeed redundant and can also be removed from the regression sets without relevant loss.

Figure 4.10 shows the split count distribution after removing LPC envelopes of order 4, 8 and 16. It can be seen that all the remaining envelope generating processes have a more uniform split

Table 4.12: LPC options experiment - Regressed Values Error - Bagging

Models		Regressed Parameters error								
		PEF	FEV ₁	FVC	Tiff	MMEF	Cust1	MEF25	MEF50	MEF75
Avg. %	All LPC	21.57	31.52	22.2353	9.98	42.86	25.14	70.90	43.32	34.21
	LPC A	20.71	32.00	24.0082	10.17	40.71	25.52	77.31	40.51	34.32
	LPC B	22.23	31.82	23.1340	9.99	41.64	24.96	77.76	41.44	34.44
Std. %	All LPC	23.86	28.80	20.2520	8.86	53.87	26.71	91.35	47.45	43.36
	LPC A	21.07	31.01	22.7291	9.03	53.50	28.09	90.68	46.51	46.70
	LPC B	22.84	28.44	22.3020	8.78	53.71	27.75	108.34	47.81	47.86

Table 4.13: LPC options experiment - Regressed Values Error - Random Forest

Models		Regressed Parameters error								
		PEF	FEV ₁	FVC	Tiff	MMEF	Cust1	MEF25	MEF50	MEF75
Avg. %	All LPC	22.59	29.23	23.85	10.11	39.66	24.35	74.73	41.64	37.10
	LPC A	21.60	29.61	24.20	9.42	41.02	25.58	70.46	42.35	34.36
	LPC B	21.85	29.96	23.77	10.01	41.81	25.23	71.84	38.93	37.26
Std. %	All LPC	22.83	25.82	21.33	9.01	49.22	25.92	96.44	46.46	47.81
	LPC A	21.15	26.95	22.95	8.54	53.96	26.99	82.33	46.74	42.50
	LPC B	20.17	25.47	24.89	8.74	53.44	28.29	90.09	40.90	47.89

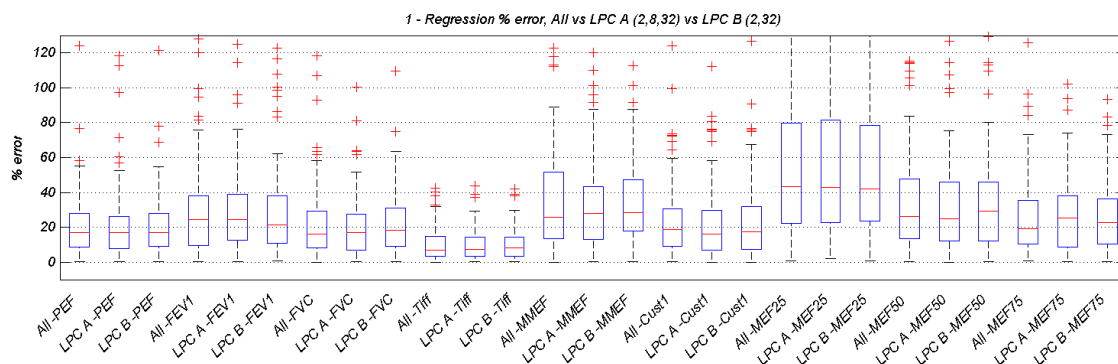


Figure 4.9: Comparison of subsets of LPC envelopes' and their influences on Bagging Regression results. Error Distribution, throughout the clinical parameters (enlargement)

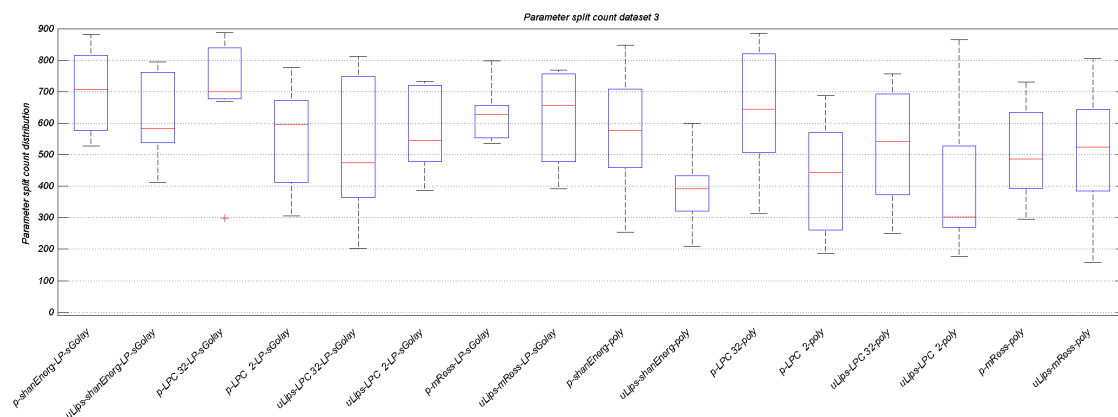


Figure 4.10: Split count distribution over the 9 parameters, for LPC set B (containing only orders 2 and 32)

count distribution than with the previous amount of methods. Reaching this point, there would be little further to do in order to remove additional methods, retaining the lowest prediction error. However, in the next sections some still relevant experiments are discussed.

4.4.2 Low Pass Filtering and Savitzky-Golay Smoothing

SpiroSmart [19] employs Savitzky-Golay smoothing in order to perfect the envelopes' form, "denoising" it. However, since their main objective is to regress the envelope, they also keep the Low Pass filtered envelopes in order to get the best of both worlds, more noisy peaks and a more "truefull" response by using LPFed envelopes and a more spirometricly realistic and robust exhalation curve when cascading the smoothing.

Until now, all curves were either smoothed with Savitzky-Golay filter or approximated by a polynomial. This experiment seeks to find if using both types of envelopes is the best option for parameter regression or if using just LP filtering is as accurate in regressing those values, which would enable to discard the smoothing from the process.

Table 4.14: LPF and Savitzky-Golay experiment - Regressed Values Error - Bagging

Models		Regressed Parameters error								
		PEF	FEV ₁	FVC	Tiff	MMEF	Cust1	MEF25	MEF50	MEF75
Avg.%	Both	20.60	31.02	23.38	9.91	41.83	24.96	76.01	40.15	36.95
	SGolay	21.64	31.21	25.46	10.15	42.72	25.50	75.33	40.47	34.10
	Just LP	21.83	31.29	24.12	9.81	43.07	24.89	76.21	43.13	34.47
Std.%	Both	21.02	28.73	23.80	8.86	53.17	27.30	108.54	44.88	45.51
	SGolay	21.14	28.65	22.92	8.75	52.97	27.82	97.54	41.05	41.53
	Just LP	20.54	28.49	25.28	9.11	55.41	27.95	101.71	48.06	43.29

Table 4.15: LPF and Savitzky-Golay experiment - Regressed Values Error - Random Forest

Models		Regressed Parameters error								
		PEF	FEV ₁	FVC	Tiff	MMEF	Cust1	MEF25	MEF50	MEF75
Avg.%	Both	21.10	30.76	24.69	9.84	43.86	24.09	71.30	43.66	33.25
	SGolay	22.49	31.00	23.95	10.06	43.50	25.10	73.71	40.89	34.66
	Just LP	21.51	29.71	23.82	10.01	43.18	25.43	70.98	43.06	34.97
Std.%	Both	20.86	29.19	24.57	9.28	51.59	26.58	86.15	46.41	41.66
	SGolay	22.61	27.91	21.55	8.65	53.59	26.81	86.78	44.23	44.89
	Just LP	20.73	26.91	22.01	8.87	53.36	28.09	83.58	51.81	47.35

Figure 4.11 and Tables 4.14 and 4.15 present the results from adding the LP filtered envelopes directly and removing the Savitzky-Golay smoothing from the process. The set in which *Both* envelope types are used does not significantly improve the parameter's prediction. On the best case, it improves the error average by 2%, which is hardly significant considering the relatively high standard deviation.

This suggests that only keeping the direct filtered envelopes should be enough for regression. This hypothesis is backed up by the results from the split count distribution.

Figure 4.12 shows the split count distribution for the set which uses LP filtered envelopes directly as well as Savitzky-Golay smoothed envelopes. This figure is particularly eloquent concerning the relative importance of the smoothed envelopes for regressing the parameters. Only the complete set of smoothed envelopes are on the bottom 35% of total split count (counting all

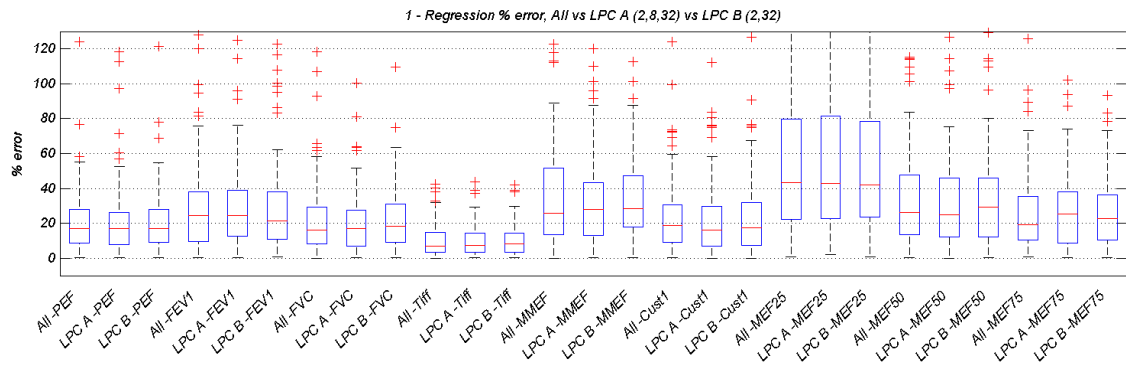


Figure 4.11: Comparison of the Low Pass filter and Savitzky-Golay envelopes' and their influences on Bagging Regression results. Error Distribution, throughout the clinical parameters (enlargement)

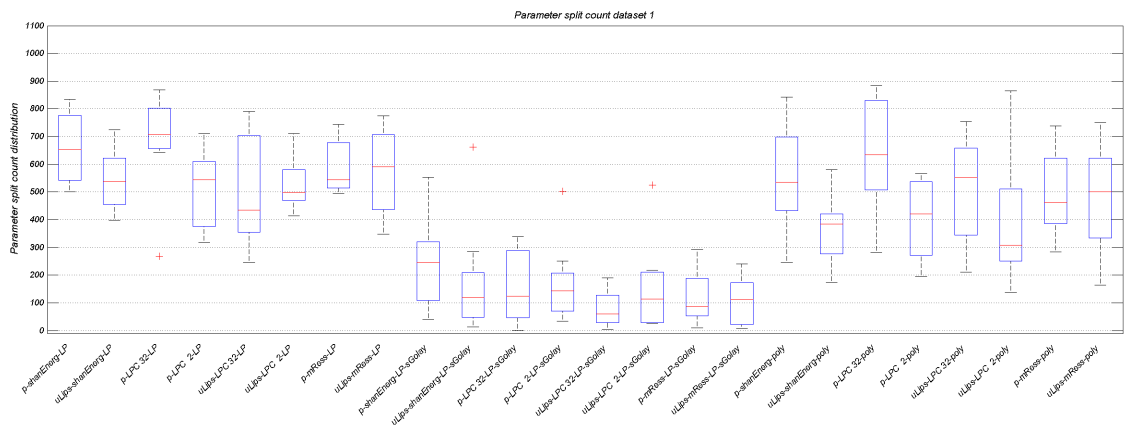


Figure 4.12: Split count distribution over the 9 parameters, for the Savitzky-Golay and Low Pass filtering options (using both)

regressed parameters' results) which makes their contribution on the distribution relatively small as well.

These results clearly indicate that the Savitzky-Golay smoothing could be removed from the process without considerable downfall. This reduces the post-processing set to just Low Pass filtering and Polynomial fitting.

4.4.3 Polynomial Fitting

For the final parameter regression experiment, the importance of the polynomial fitting is going to be assessed. In past projects [6, 13], the envelope regression was based on polynomial approximations and had relatively high success on regressing the lung function parameters on their datasets.

This experiment consists of either removing the polynomial approximation from the parameter regression set or the Low Pass filtering. As control sample, the method set using both post-processing methods shall be compared as well.

Table 4.16: Polynomial Fitting experiment - Regressed Values Error - Bagging

Models		Regressed Parameters error								
		PEF	FEV ₁	FVC	Tiff	MMEF	Cust1	MEF25	MEF50	MEF75
Avg.%	Both	22.51	31.43	23.18	10.12	43.54	24.60	75.77	40.77	34.63
	No Poly	21.33	32.21	22.60	10.02	41.40	26.95	70.50	42.65	34.12
	No LPF	21.41	32.45	22.26	9.53	47.86	24.19	76.34	43.47	36.89
Std.%	Both	20.43	28.29	21.94	9.00	55.31	28.24	100.01	43.53	46.20
	No Poly	18.82	26.12	21.93	8.49	47.30	25.51	101.06	47.37	42.05
	No LPF	24.85	31.94	22.23	9.22	58.87	30.07	93.66	48.77	46.15

Table 4.17: Polynomial Fitting experiment - Regressed Values Error - Random Forest

Models		Regressed Parameters error								
		PEF	FEV ₁	FVC	Tiff	MMEF	Cust1	MEF25	MEF50	MEF75
Avg.%	Both	22.59	31.01	23.27	10.16	43.40	24.21	70.97	43.72	34.47
	No Poly	21.40	30.41	24.34	10.02	41.33	26.97	73.60	41.93	34.32
	No LPF	22.12	30.76	25.12	9.58	47.94	24.60	69.97	41.10	35.39
Std.%	Both	20.79	27.63	23.28	9.09	54.47	26.36	81.22	45.43	43.10
	No Poly	19.81	27.09	21.52	8.81	48.18	26.23	90.88	47.80	43.84
	No LPF	24.52	30.19	24.46	8.66	58.75	27.30	83.26	44.06	44.40

Figure 4.13 and Tables 4.16 and 4.17 present the parameter regression results from the removal of either post-processing method. The errors' average do not vary significantly between the three sets and thus, one of them can be removed from the process without compromising the regression results. The choice must fall again on method complexity and performance issues.

Although polynomial fitting provides a smoother and, apparently, more robust curve, it requires more computational resources than a Butterworth filter. The polynomial approximation can be done in an iterative fashion using Least Squares or using a Gauss-Jordan elimination as done in [6]. In either method, each curve data point or error, is taken to the power of 2 or above (in the case of Gauss-Jordan), and may be computed more than once (on the iterative process).

A Butterworth low pass filter, however, only requires to be computed once, the 2 coefficients (for order 2, as the one used) and a few additions, over all curve points.

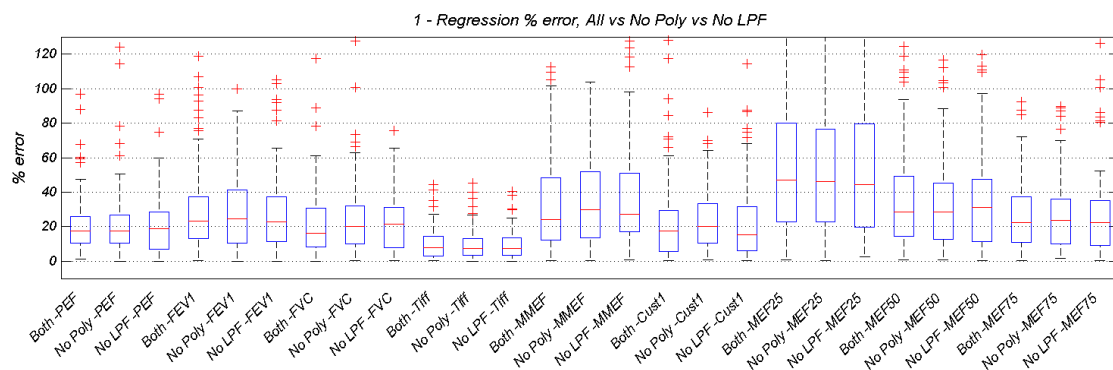


Figure 4.13: Comparison of the Low Pass filter and Polynomial Fitting envelopes' and their influences on Bagging Regression results. Error Distribution, throughout the clinical parameters (enlargement)

In addition to this performance argument, a physician could obtain more relevant information from the un-smoothed LPF curve than an unrealistic polynomial fit, by merely observing the curve's slope, as depicted on Figure 4.14.

Therefore, it was decided to remove the polynomial approximation, leaving only LPF on the post-processing methods. However, as a classification baseline, and curiosity as to the extent of its robustness, both regressed datasets (containing either polynomial fitted or LP filtered envelopes) were initially tested.

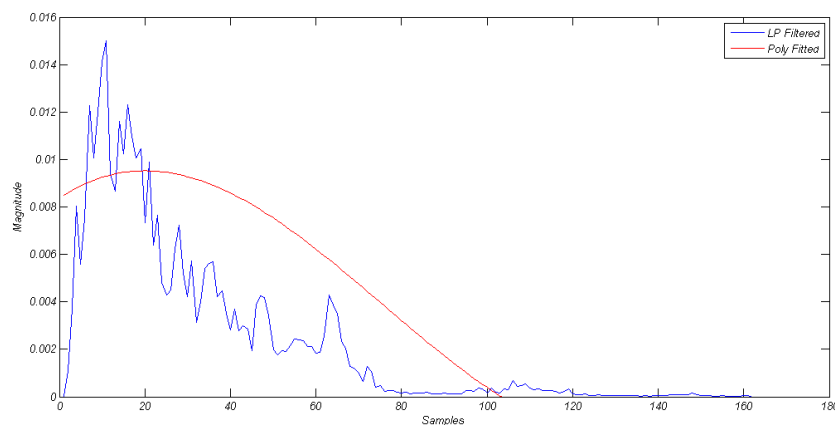


Figure 4.14: Comparison between a filtered envelope and a polynomial fitted envelope

4.4.4 Final Regression Model

After all these experiments, the Signal Processing methods that were employed in the final regression model are the following:

- Pre-processing: Flow Conversion Model,

- Envelope generation:
 - Shannon Energy,
 - Linear Predictive Coding, orders 2 and 32,
 - Mean of Resonances,
- Post-processing: Low Pass Butterworth Filter.

Figure 4.15 presents a diagram of the Signal Processing system.

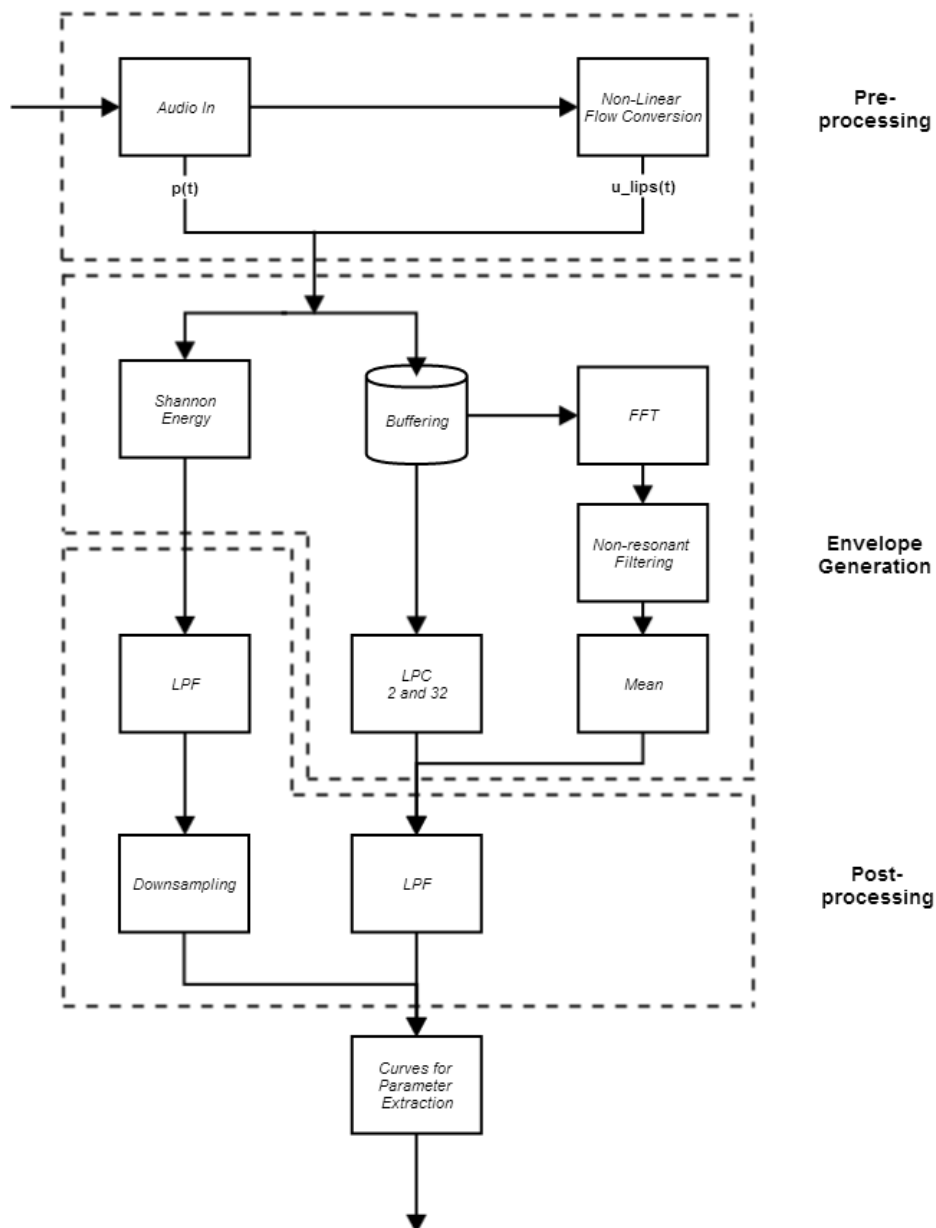


Figure 4.15: Signal Processing stage Diagram

4.5 Two Label Classification

Due to an severely uneven distribution of labels and for simplicity's sake, the classification process was first devised as a normal against abnormal classification problem, ignoring both the distinction between obstruction and restriction and the several degrees of the pathologies.

The linear learners/classifiers should easily discard the degrees of the pathologies, without compromising much of the classification. The worst case scenario would be not segmenting the data at the place that proves to guarantee the best margin. However, since obstruction and restriction (and mixed abnormalities) have significantly different parameter patterns, probably tree based methods would surpass SVM methods since the first are based on several hyperplane splits, tending to overfit, while the later generates a single hypercut/hypersplit and would probably need a kernel transformation, for a relatively small feature space. Either way, the following classifiers were experimented upon:

- Single Classification Tree (Baseline)
- Tree Bagging
- Tree Bagging using the regression targets as dataset (*GTruth*)
- Random Forest
- AdaBoost Trees
- μ -SVM / soft margin, with several kernel functions
- Naïve Bayes

4.5.1 Tree Ensemble Setup

Since ensemble methods generally have an optimal number of learners for reducing the decision error, a series of 5 runs of 5-fold, tree bagging learning and prediction process was made. The number of trees to use on the ensemble methods was decided based on the average misclassification count of each ensemble set. Table 4.18 shows the results of those 5 simulations for the various numbers of trees.

All the classification errors are quite similar. The 70 trees set gave the second smallest average misclassification count, although only slightly less than the best option. In this case, choosing the second best can also reduce significantly the time spent teaching the learner how to classify and providing a prediction. For Boosted methods, however, we chose to use 10 iterations/trees in order to avoid overfitting.

4.5.2 Classification with the 9 clinical parameters

Once the tree number was selected for the ensemble methods, the regression output was used to train the several classification methods and test them, in a 5 fold sense. Tables 4.19 and 4.20 the

Table 4.18: Classification percentage error - Tree Bagging test

Number of Trees	Average %
20	24.35
30	24.35
40	22.37
50	22.17
60	22.97
70	21.38
80	23.16
90	24.15
100	23.56
110	22.77
120	24.95
130	23.96
140	25.14
150	22.17
160	21.18
170	22.17
180	22.97
190	23.36

misclassification results, as percentage error, precision and recall. Precision is defined as equation 4.1a and recall as 4.1b.

$$\text{Precision} = \frac{tp}{(tp + fp)} \quad (4.1a)$$

$$\text{Recall} = \frac{tp}{(tp + fn)} \quad (4.1b)$$

where tp refers to the true positive count (Normal labeled observations predicted as Normal class), fp refers to the false positive count (Abnormal labeled observations predicted as Normal class) and fn refers to the false negative count (Normal labeled observations predicted as Abnormal class).

Although the method of choice for post-processing was the Low Pass filter, the regressed dataset using the polynomial fitting was also tried for the classification models.

The *SingleTree-Bag* is a classification model based on a single tree which uses all features, simultaneously, while trying to find the optimal splits, while *SingleTree-RF* is the same model with the exception of only considering a portion of the features (\sqrt{D}) in each split, like in Random Forest method. *AdaBoostMI Trees* is the AdaBoost method applied to decision tree classifiers, for a binary classification problem. The several *SVM* models use different kernels, though all their coefficients and internal parameters were left as LIBSVM's default values.

In addition to the regressed parameter trained classification models, three other classification models were implemented.

These *GTruth* models, which stands for Ground Truth, are classifiers trained and tested directly with the regression targets instead of the regressed values, in a 5 fold sense. This model also acts as a baseline for both the machine learning models' performance as for the regressed values results. Since there are less regression targets than recordings (multiple recordings per patient) this

particular dataset is only constituted by 61 patients/values, from which 47 are labeled as Normal and 14 are labeled as Abnormal lung function.

Table 4.19: Classification percentage error, precision and recall for Normal label, on the polynomial fitting set (101 observations, 61 patients)

Method	Error	Precision	Recall
SingleTree - Bag	33.66	76.05	76.05
SingleTree - RF	35.64	74.64	74.64
Bagging Trees	25.74	80.00	84.50
Random Forest	19.80	84.93	87.32
AdaBoostM1 Trees	22.77	82.43	85.91
Naïve Bayes	26.73	87.93	71.83
SVM - Linear	22.77	82.43	85.91
SVM - Poly-2	29.70	70.29	100
SVM - Poly-3	29.70	70.29	100
SVM - RBF	34.65	69.56	90.14
SVM - Sigmoid	29.70	70.29	100
SVM - OneClass	30.69	70.40	97.18
GTruth - Bag Trees	24.59	82.00	87.23
GTruth - RF Trees	18.03	84.61	93.61
GTruth - AdaBoostM1	21.31	85.41	87.23

Table 4.20: Classification percentage error, precision and recall for Normal label, on the LPF filtering set (101 observations, 61 patients)

Method	Error	Precision	Recall
SingleTree - Bag	34.65	73.68	78.87
SingleTree - RF	40.59	71.42	70.42
Bagging Trees	29.70	78.08	80.28
Random Forest	24.75	81.08	84.50
AdaBoostM1 Trees	28.71	78.37	81.69
Naïve Bayes	32.67	80.64	70.42
SVM - Linear	28.71	76.92	84.50
SVM - Poly-2	29.70	70.29	100
SVM - Poly-3	29.70	70.29	100
SVM - RBF	32.67	70.21	92.95
SVM - Sigmoid	29.70	70.29	100
SVM - OneClass	32.67	69.79	94.36
GTruth - Bag Trees	26.22	81.63	85.10
GTruth - RF Trees	29.50	80.85	80.85
GTruth - AdaBoostM1	27.86	82.60	80.85

As expected from a training set based on very noisy and poorly repeatable recordings, the misclassifications count is considerable. Among all the classifiers, the average error is 29.8, for the polynomial fitting dataset, and 30.5, for the LPF dataset, in 101. Additionally, as previously mentioned, the abnormal classification also includes values from three pathologies with different parameter patterns, lung Obstruction, Restriction and Mixed abnormalities.

Nevertheless, the proprieties of each classifier can be studied and compared with the expected results and with each other. As mentioned in Section 3.4.2, generally, ensemble methods tend to perform better than single learner methods. This can be observed by comparing the ensemble methods results (*Bagging Trees*, *Random Forest*, *AdaBoostM1 Trees*) to non-ensemble classifiers' (*SingleTree - Bag* and *RF* and even *SVM*, although with less significance).

For exactly the same folding distribution, Random Forest performed a little better than Ad-aBoost, corroborating the fact that both methods obtain very similar results and both best normal,

full feature, bagging methods. Additionally, *Linear SVM* also performed relatively well, being on par with AdaBoost.

Naïve Bayes almost sets a frontier between the best and worst methods. Although, unlike every other tested method, it tends to mistake normal labels for abnormal ones, having therefore, a significantly lower recall than precision. This might be due to the scatter pattern of the abnormal class and the classifier being hyperspatial radial based.

Kernel based SVM models mostly classified all samples as normal, which might indicate that the present conditions might not be the best, or that using a kernel is unnecessary.

GTruth models had fairly good results comparing with regression valued classification models. This was not unexpected since the spirometers' measurement errors have to be within 6% [20] and they also are the regression target values, which may propagate the error to the regressed values. Even between the *GTruth* models, the difference in performance of the three methods is noticeable and consistent with difference in performance regressed values' ensemble methods.

Despite having initially chosen the LPF only regression dataset, these results show that, indeed, the robustness of the polynomial approximation is a property that should be kept. All tree related methods, as well as Linear SVM, suggest that the dataset to use should, instead, be the one with the polynomial fitting. Therefore, hereafter, all classification tests were conducted using this last post-processing method. Figure 4.16 presents the new and final signal processing stage of the system.

4.5.3 Classification with the 4 most used clinical parameters

The same experiment was conducted using only the 4 most popular parameters for lung function evaluation, PEF, FEV₁, FVC and Tiff. The result are shown on Table 4.21.

Table 4.21: Influence of Height and Age on the Classification percentage error, precision and recall for Normal labels (101 observations, 61 patients)

Method	Error	Precision	Recall
SingleTree - Bag	39.60	70.12	76.05
SingleTree - RF	35.64	75.36	73.23
Bagging Trees	32.67	73.75	83.09
Random Forest	30.69	77.77	78.87
AdaBoostM1 Trees	33.66	73.41	81.69
Naïve Bayes	30.69	82.25	71.83
SVM - Linear	29.70	70.29	100
SVM - Poly-2	29.70	70.29	100
SVM - Poly-3	29.70	70.29	100
SVM - RBF	33.66	69.07	94.36
SVM - Sigmoid	29.70	70.29	100
SVM - OneClass	37.62	77.04	66.19
GTruth - Bag Trees	22.95	82.35	89.36
GTruth - RF Trees	16.39	84.90	95.74
GTruth - AdaBoostM1	21.31	86.95	85.10

In general, all methods had worse predictions. As expected, ensemble Random Forest becomes worse due to the reduction of both the total number of parameters and the number of splittable parameters over each node. However, kernel based SVMs results did not change significantly, except for *OneClass* which became around 7% worse.

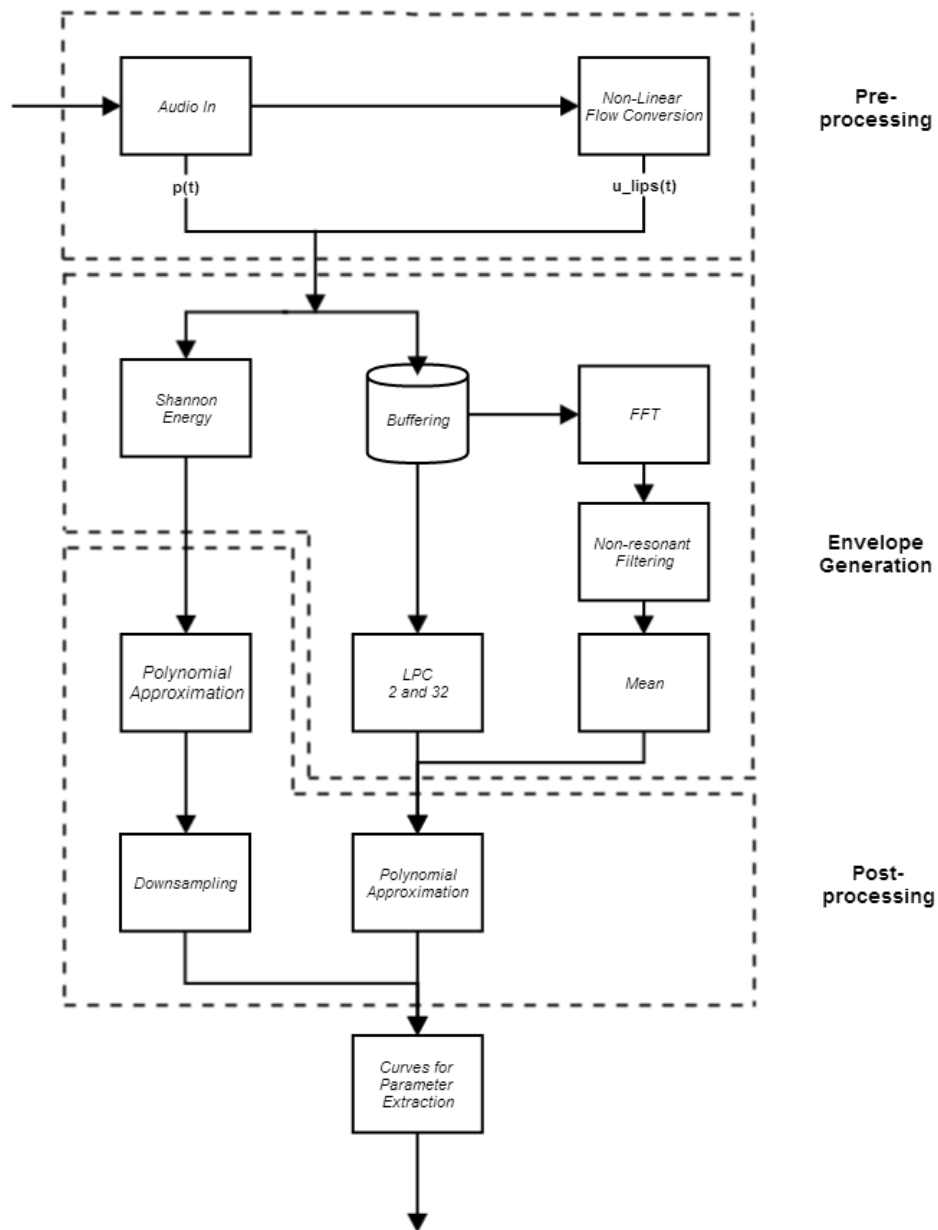


Figure 4.16: Signal Processing stage Final Diagram

GTruth models, on the other hand, improved a little over the previous results. Since this dataset is less error prone, this suggests that the parameters removed are error inducing values.

4.5.4 Influence of Height and Age in Classification feature space

For the final experiments, the patients' height was included on the training and testing datasets, then the patient's age was also included. For the last dataset, the least relevant parameters were removed, leaving only PEF, FEV₁, FVC, Tiff, height and age. The results of the three learning and classification processes are shown on Table 4.22.

Table 4.22: Classification percentage error, precision and recall for Normal label, using using different sets of parameters (101 observations, 61 patients)

Method	Plus Height			Plus Height and Age			4 most popular, Plus Height and Age		
	Error	Precision	Recall	Error	Precision	Recall	Error	Precision	Recall
SingleTree - Bag	31.68	80.00	73.23	17.82	92.06	81.69	20.79	85.71	84.50
SingleTree - RF	33.66	77.61	73.23	29.70	78.08	80.28	29.70	78.87	78.87
Bagging Trees	17.82	87.32	87.32	11.88	88.31	95.77	9.90	91.78	94.36
Random Forest	19.80	84.00	88.73	15.84	85.71	92.95	7.92	90.90	98.59
AdaBoostM1 Trees	21.78	83.56	85.91	16.83	85.52	91.54	9.90	91.78	94.36
Naïve Bayes	24.75	87.09	76.05	23.76	88.52	76.05	16.83	86.48	90.14
SVM - Linear	17.82	86.30	88.73	9.90	92.95	92.95	7.92	94.36	94.36
SVM - Poly-2	29.70	70.29	100	29.70	70.29	100	29.70	70.29	100
SVM - Poly-3	29.70	70.29	100	29.70	70.29	100	29.70	70.29	100
SVM - RBF	30.69	70.00	98.59	29.70	70.29	100	20.79	77.17	100
SVM - Sigmoid	29.70	70.29	100	29.70	70.29	100	29.70	70.29	100
SVM - OneClass	29.70	70.29	100	29.70	70.29	100	30.69	70.00	98.59
GTruth - Bag Trees	22.95	82.35	89.36	21.31	82.69	91.48	21.31	85.41	87.23
GTruth - RF Trees	21.31	84.00	89.36	19.67	85.71	89.36	21.31	86.95	85.10
GTruth - AdaBoostM1	27.86	80.00	85.10	26.22	82.97	82.97	24.59	84.78	82.97

On the first parameter set, there seems to be a reduction of the misclassification count on the 4 most prominent classifiers. In fact, Bagged Trees and Linear SVM outperformed both Random Forest and AdaBoost, even if only slightly.

When Age is introduced in the parameter set, the results of those classifiers are even better. This is somewhat expected considering that both height and age are relevant anthropometric parameters. These parameters are usually used by physicians to evaluate if the patient's PEF, FEV₁ and FVC are in the normal values range for their physiognomy or not.

The results from the third set parameter further reinforce these parameters' importance. By leaving only the 4 most medically relevant parameters, height and age, the models were enabled to focus on the most valuable information. In fact, most of the parameters removed, MEF_x and MMEF, were parameters that are considered very unstable and extremely effort dependent. Therefore, they most likely only introduced noise on the classifiers' choices.

Concerning the *GTruth* models, each step improved the AdaBoost results, which also supports the MMEF and MEF_x statement. On the other hand, Bagging and Random Forest results stayed more or less the same, which might indicate that the 180 trees, in the ensemble, might be inducing overfitting, on this dataset. Either way, these three models do not improve to the extent of the regressed valued classification models, which might indicate that those models are promising or that *GTruth* dataset is just reduced and less variable.

Summing up the results of the two label classification problem, the best parameter combination for the classification model would be PEF, FEV₁, FVC, FEV₁/FVC (Tiff), height and age. These results however, do not enable a clear decision concerning the best classification model to use. If the simplicity of the implementation is the major issue, probably either Bagging or Random Forest method should be used. Although, once the boosted method generates the trees and their voting weight, it becomes as simple as the other two. Likewise, Linear SVM trained classifier would be fairly simple to implement however, if an online method was necessary to be updated, than it would not be the case.

4.6 Multiple Label Classification

In this section, some experiments of the multiple label classification problem are discussed. While the last section approached the prediction into normal or abnormal lung function, these experiments show how the learners deal with the 4 different classes: normal, obstruction, restriction and mixed abnormalities.

The last section's results shed some light on what to expect from the classifiers trained with this regressed dataset. However, in this classification task, the considerably smaller sub-dataset (Abnormal label) is scattered among the 3 abnormal classifications, which reduces significantly the number observations per class. This is expected to increase the misclassification count, if not for predicting abnormal class observations as another, nearby, abnormal class.

Also, due to the limited number of observations, some changes had to be made to the normal 5-fold distribution. For instance, since there are only two lung restriction observations, each training set had to include one of the observations, while the testing set had to include the other.

The Naïve Bayes classifier could not be trained with all labels since in each class, at least two observations must be present. This condition could not be fulfilled due to the aforementioned training and testing scheme. As an alternative, for this classifier only, the restriction label observations were discarded.

Additionally, since the mixed abnormality observations came from the same patient, both height and age are the same. This made it impossible for the Naïve Bayes classifier to find the variance for those features and, therefore, the Mixed class observations could not be used. To circumvent that issue, for the subsequent experiments, both height and age, of all observations, were added a random amount (from an uniform distribution), which ranged from 0 to 1/1000 (centimeters and years, respectively). This enabled the use of the Mixed abnormality observations and changed minimally the original values.

4.6.1 Classification with the 9 clinical parameters

In order to establish a comparison baseline, the first multi-class experiment uses the same feature space as the first dual-class experiment. This way, the classification issues from both problems (dual and multi-class) can be addressed and a starting point for the subsequent multi-class experiments is provided.

Table 4.23: Linearized Confusion Matrices for the Classification with the 9 clinical regressed parameters

Methods	Confusion Matrix																TErr
	Predicted as Nm				Predicted as Ob				Predicted as Rs				Predicted as Mx				
	Nm	Ob	Rs	Mx	Nm	Ob	Rs	Mx	Nm	Ob	Rs	Mx	Nm	Ob	Rs	Mx	
SingleTree - Bag	62	13	2	3	5	8	0	1	2	0	0	1	2	2	0	0	31
SingleTree - RF	51	10	2	3	13	9	0	1	3	0	0	0	4	4	0	1	40
Bagging Trees	61	13	2	2	10	10	0	2	0	0	0	0	0	0	0	1	29
Random Forest	64	12	2	2	7	11	0	2	0	0	0	0	0	0	0	1	25
AdaBoostM2	66	15	2	2	5	7	0	2	0	0	0	0	0	1	0	1	27
Trees																	
Naïve Bayes	53	5	-	1	18	18	-	4	-	-	-	-	0	0	-	0	28
SVM - Linear	61	10	2	4	8	13	0	1	2	0	0	0	0	0	0	0	27
SVM - Poly-2	71	23	2	5	0	0	0	0	0	0	0	0	0	0	0	0	30
SVM - Poly-3	71	23	2	5	0	0	0	0	0	0	0	0	0	0	0	0	30
SVM - RBF	67	22	2	3	4	1	0	2	0	0	0	0	0	0	0	0	33
SVM - Sigmoid	71	23	2	5	0	0	0	0	0	0	0	0	0	0	0	0	30

Table 4.23 shows the linearized confusion matrices of each method. *Nm* refers to the Normal lung function label, *Ob* to Obstructive abnormality label, *Rs* to Restrictive abnormality label and *Mx* to Mixed abnormality label. The *TErr* column indicates the total misclassification count across all labels.

As expected, the total number of misclassifications increased when compared with the experiment in Section 4.5.2. Not only a misclassification is considered when an abnormal label is mistaken for a normal one, but also when abnormal labels are mistaken between themselves.

The restriction labeled observations were predicted as normal label by all classifiers, which indicates that either the regressed values are far off the expected values for a lung restriction pattern, which might have crippled the learner, or that indeed learning from one observation is considerably difficult.

Like in the experiment mentioned above, SVM classifiers with non-linear kernels maintain about the same misclassification count and false Normal prediction.

Among the best classifiers are still Random Forests, AdaBoost (in the multi-class version) and Linear SVM. Naïve Bayes also preformed relatively well, although it had less two observations to train and classify, as well as one less label to consider. Finally, as in the original, two label prediction problem, the *SingleTree* methods perform worse than ensemble methods.

4.6.2 Classification with the 4 most popular clinical parameters

As in Section 4.5.3, the second experiment required that the less clinically relevant parameters would be discarded, maintaining only the 4 most popular: PEF, FEV₁, FVC and FEV₁/FVC (Tiff). From the previous results, it was expected that the removal of these parameters would also reduce a little of the overfitting effort which implies that misclassification would increase.

Table 4.24 shows the linearized confusion matrices concerning the current experiment. As has happened on the equivalent experiment for the two label classification, just by removing the parameter subset, the classification results generally worsened. The best classifier models increased their total misclassification count by an amount from 2 to 8, becoming on par with the previously

Table 4.24: Linearized Confusion Matrices for the Classification with the 9 clinical regressed parameters plus height

Methods	Confusion Matrix																TErr
	Predicted as Nm				Predicted as Ob				Predicted as Rs				Predicted as Mx				
	Nm	Ob	Rs	Mx	Nm	Ob	Rs	Mx	Nm	Ob	Rs	Mx	Nm	Ob	Rs	Mx	
SingleTree - Bag	49	17	1	4	14	4	1	1	0	0	0	0	8	2	0	0	48
SingleTree - RF	49	16	0	3	21	6	2	2	1	0	0	0	0	1	0	0	46
Bagging Trees	59	12	1	3	12	10	1	1	0	0	0	0	0	1	0	1	31
Random Forest	61	16	1	4	10	7	1	1	0	0	0	0	0	0	0	0	33
AdaBoostM2 Trees	62	17	2	3	9	6	0	2	0	0	0	0	0	0	0	0	33
Naïve Bayes	57	15	-	2	13	7	-	3	-	-	-	-	1	1	-	0	35
SVM - Linear	71	23	2	5	0	0	0	0	0	0	0	0	0	0	0	0	30
SVM - Poly-2	71	23	2	5	0	0	0	0	0	0	0	0	0	0	0	0	30
SVM - Poly-3	71	23	2	5	0	0	0	0	0	0	0	0	0	0	0	0	30
SVM - RBF	70	23	2	5	1	0	0	0	0	0	0	0	0	0	0	0	31
SVM - Sigmoid	71	23	2	5	0	0	0	0	0	0	0	0	0	0	0	0	30

worse methods such as SVMs with non-linear kernels. However, unlike those, the errors are distributed among the other classes, i.e., a significant portion of abnormal labels is not mistakenly predicted as normal. On the other hand, the problem shifted towards a normal versus obstructive abnormality dichotomy, which is somewhat similar to the normal versus abnormal, two label problem. Kernel based SVM models more or less maintained their misclassification distribution.

4.6.3 Influence of Height and Age in Classification feature space

The following experiment aims to find the importance of the patients' height and age when trying to classify the lung function based on the regressed parameters. As with the first experiment on Section 4.5.4, the 9 parameters were kept on the feature space and both height and age were also included. The results from this experiment are shown on Table 4.25.

Table 4.25: Linearized Confusion Matrices for the Classification with the 9 clinical regressed parameters, plus height and age

Methods	Confusion Matrix																TErr
	Predicted as Nm				Predicted as Ob				Predicted as Rs				Predicted as Mx				
	Nm	Ob	Rs	Mx	Nm	Ob	Rs	Mx	Nm	Ob	Rs	Mx	Nm	Ob	Rs	Mx	
SingleTree - Bag	58	7	2	2	8	12	0	1	1	2	0	0	4	2	0	2	29
SingleTree - RF	63	12	1	1	6	8	0	1	2	0	1	0	0	3	0	3	26
Bagging Trees	68	9	2	1	3	13	0	3	0	0	0	0	0	1	0	1	19
Random Forest	65	10	2	2	6	13	0	3	0	0	0	0	0	0	0	0	23
AdaBoostM2 Trees	65	10	2	3	5	12	0	2	0	0	0	0	1	1	0	0	24
Naïve Bayes	54	3	-	1	17	20	-	0	-	-	-	-	0	0	-	4	21
SVM - Linear	68	3	2	1	1	17	0	2	1	0	0	0	1	3	0	2	14
SVM - Poly-2	71	23	2	5	0	0	0	0	0	0	0	0	0	0	0	0	30
SVM - Poly-3	71	23	2	5	0	0	0	0	0	0	0	0	0	0	0	0	30
SVM - RBF	71	23	2	5	0	0	0	0	0	0	0	0	0	0	0	0	30
SVM - Sigmoid	71	23	2	5	0	0	0	0	0	0	0	0	0	0	0	0	30

These results present a slight, overall, improvement in comparison with the original ones, in Section 4.6.1. The ensemble methods continued among the best to classify this dataset, although, Linear SVM outperformed all the others, in this case. The increased dimensionality might have provided a more valuable hyperplane split (or in this case splits). Both Normal and Obstruction

label classification improved slightly, and in the case of non-ensemble methods, so did the Mixed abnormality label.

In this experiment, the Naïve Bayes performed somewhat better than the original multi-class experiment and could correctly classify almost all of the mixed abnormalities. On the other hand, it did not improve much on the obstruction and normal labels. This might indicate that the single change, with the mixed label observations as the main target, was what caused the great improvement. This seems to be a case of severe overfitting. In fact, this hypothesis is supported by the results from the other methods: *SingleTree* (*Bag* and *FR*) were among the best in discerning mixed abnormality labels for this parameter set because the methods themselves tend to overfit. As consequence, provided new mixed abnormal lung recordings were gathered, these methods would not classify as well. The ensemble methods, however, had a greater difficulty at this task, which also contributes to the validity of the previous statements. Kernel based SVM models maintained their misclassification distribution.

4.6.4 Classification with the 4 most popular clinical parameters plus Height and Age

For the last experiment, the dataset's feature space included the 4 clinical parameters, as in the dataset in Section 4.6.2, as well as the patients' height and age. Table 4.26 presents the results of the experiment.

Table 4.26: Linearized Confusion Matrices for the Classification with the 4 most popular clinical regressed parameters, plus height and age, for the Polynomial dataset

Methods	Confusion Matrix																TErr
	Predicted as Nm				Predicted as Ob				Predicted as Rs				Predicted as Mx				
	Nm	Ob	Rs	Mx	Nm	Ob	Rs	Mx	Nm	Ob	Rs	Mx	Nm	Ob	Rs	Mx	
SingleTree - Bag	59	4	1	1	8	16	1	1	3	0	0	0	1	3	0	3	23
SingleTree - RF	60	8	1	2	8	14	1	1	1	0	0	0	2	1	0	2	25
Bagging Trees	66	2	2	0	5	21	0	1	0	0	0	0	0	0	0	4	10
Random Forest	69	4	2	1	2	19	0	1	0	0	0	0	0	0	0	3	10
AdaBoostM2	70	5	2	1	1	17	0	1	0	0	0	0	0	1	0	3	11
Trees	67	7	-	0	4	16	-	0	-	-	-	-	0	0	-	5	11
Naïve Bayes	69	6	2	0	1	17	0	4	0	0	0	0	1	0	0	1	14
SVM - Linear	71	23	2	5	0	0	0	0	0	0	0	0	0	0	0	0	30
SVM - Poly-2	71	23	2	5	0	0	0	0	0	0	0	0	0	0	0	0	30
SVM - Poly-3	71	23	2	5	0	0	0	0	0	0	0	0	0	0	0	0	30
SVM - RBF	71	14	2	5	0	9	0	0	0	0	0	0	0	0	0	0	21
SVM - Sigmoid	71	23	2	5	0	0	0	0	0	0	0	0	0	0	0	0	30

The improvement from the previous experiment mainly consisted of the misclassification decrease of the obstruction label observations. In addition, the mixed abnormality observations were almost all correctly classified both in ensemble and on non-ensemble methods, disregarding the ones that have a considerable misclassification rate such as SVMs with non-linear kernels.

These results, once again, support the way physicians conduct the lung function evaluation, accounting PEF, FEV₁, FVC, FEV₁/FVC and the patients' anthropometric parameters height and age and disregarding, most of the times, the MMEF and MEF_x values due to their variability. In fact, since the exhaled volumes (in the first second and total) are only relevant when comparing

with height and age, only index-like parameters such as FEV₁/FVC would amount to any reasonable classification, when lacking the anthropometric parameters.

Considering the relatively small dataset and the somewhat harsh recording environment, the best methods managed to produce classification errors under 10%.

The results of the multiple label classification problem were similar to the binary label classification problem. In fact, even after scattering the already small number of abnormal observations, the methods were on par with their binary model counterparts. There is also the issue of close mistakes such as predicting mixed abnormality when it should be obstruction, which will not be approached in this study due to the small contribution to the abnormal lung function recorded population. Based on the results from these last multi-label classification experiments and the first binary label ones, the best methods for classifying a dataset similar to this would be the simpler implementation, ensemble methods, namely Tree Bagging or Random Forests.

Table 4.27: Linearized Confusion Matrices for the Classification with the 4 most popular clinical regressed parameters, plus height and age, for the LPF dataset

Methods	Confusion Matrix																TErr
	Predicted as Nm				Predicted as Ob				Predicted as Rs				Predicted as Mx				
	Nm	Ob	Rs	Mx	Nm	Ob	Rs	Mx	Nm	Ob	Rs	Mx	Nm	Ob	Rs	Mx	
SingleTree - Bag	61	7	2	1	8	15	0	2	0	0	0	1	2	1	0	1	24
SingleTree - RF	59	8	1	2	8	11	1	1	0	0	0	0	4	4	0	2	29
Bagging Trees	66	7	2	0	5	15	0	2	0	0	0	0	0	1	0	3	17
Random Forest	68	8	2	0	3	14	0	1	0	0	0	0	0	1	0	4	15
AdaBoostM2 Trees	66	9	2	0	5	14	0	2	0	0	0	1	0	0	0	2	19
Naïve Bayes	69	12	0	2	2	11	0	0	0	0	2	0	0	0	0	3	16
SVM - Linear	68	4	2	1	3	19	0	3	0	0	0	0	0	0	0	1	13
SVM - Poly-2	71	23	2	5	0	0	0	0	0	0	0	0	0	0	0	0	30
SVM - Poly-3	71	23	2	5	0	0	0	0	0	0	0	0	0	0	0	0	30
SVM - RBF	71	19	2	5	0	4	0	0	0	0	0	0	0	0	0	0	26
SVM - Sigmoid	71	23	2	5	0	0	0	0	0	0	0	0	0	0	0	0	30

Table 4.27 presents the same analysis as on Table 4.26, using, however, the LPF dataset. This enables a final comparison between the last two signal processing methods discussed, coming full circle, after the initial classification experiment on Section 4.5.2. As found on that experiment, the models trained and tested on the polynomial dataset overperform those trained and tested with the LPF dataset. All the methods that gave better results on the original experiment from Section 4.5.2, continue to do so, in the same relative proportion. These results further reinforce the change of heart, concerning the post-processing methods, on Section 4.5.2.

4.7 Final Prototype

In this chapter several experiments concerning learning models of the envelope parameter set were made.

First, the extracted parameters were regressed and the best combination of simplicity and accurate predictability of signal processing portion was made. Figure 4.16 shows that final signal processing architecture. The regression experiments also made it possible to choose Random Forests as the method to use due to the advantages over normal Bagging: higher robustness of the

classification and higher speed of learning, for the same regression error. Also, by regressing the several extracted parameters it was possible to reduce the feature space to 9 parameters.

The classification experiments provided an insightful view on the relative importance of the feature space, confirmed that MMEF and MEF_x were error inducing parameters and that, non-regressed, patient anthropometric parameters are important for the classification modeling of the problem. This enabled to further reduce the feature space to 6 parameters. Finally, although the binary class models had slightly better results than the multiple class models, it was verified that not only is possible to classify the several lung abnormalities as it would have a relatively good accuracy, providing more data was available for training and testing. Again, for performance and simplicity issues, the Random Forest ensemble approach would be the best, for the present multiple classification problem.

Table 4.28 presents the results from this and other related projects, for the parameter regression task. As shown, this project did not perform very well, regression-wise, even with a larger number of patients. However, the dataset used contained, approximately, an average of 1.6 recordings per patient, which hardly constitutes a solid dataset, with relatively poor repeatability standards. Additionally, all the dataset was tested using 5-fold cross validation. Like in mCOPD's work, recordings were both gathered in controlled and uncontrolled environments, although, the learning and testing process did not make a distinction from which set it would choose the recordings from. Additionally, the uncontrolled environment recordings outnumbered the controlled environment ones by 81 to 20.

SpiroApp [6], although not providing the average clinical parameters' regression error, states that the error generally did not exceed 0.5 liters. They also only conducted one test on one COPD and one Asthma patients each.

Table 4.28: Clinical parameters' results comparison

Methods	Parameters' mean error				N° of patients
	PEF	FEV ₁	FVC	FEV ₁ /FVC	
SpiroSmart [19]	6.3	4.8	5.2	4.0	52
mCOPD [13]	-	3.6	6.5	3.9	40
SpiroSS ²	21.3	32.2	22.6	10.0	61

In terms of classification, all of the projects mentioned earlier either do not offer any system or do provide such information by non-learning, rule based, guidelines. This dissertation's system provides two classification models: one for the Normal versus Abnormal classification problem and a second for the decision into Normal, Obstructive, Restrictive and Mixed abnormalities labels. The models with the best results attained around 8% and 10% misclassification rate, respectively.

²SpiroSS was the unofficially designation of this dissertation's project.

Chapter 5

Conclusion and Discussion

5.1 Conclusion

The focus of this project was to evaluate several sets of signal processing and machine learning methods, in order to find a suitable, efficient and accurate architecture for a smartphone app that measures and classifies lung function, through microphone recordings.

The final signal processing stage of the prototype is divided in three components: pre-processing, envelope generation and post-processing. On the pre-processing portion, the audio signal is fed to a non-linear model that converts the sound pressure into airflow. Both processed and unprocessed signals become inputs of the envelope generation functions. These are the Shannon Energy, Linear Predictive Coding (LPC), of orders 2 and 32, and the Mean of Resonances. After the envelopes are produced, they are approximated by a polynomial of the fourth degree. Then the envelopes are extracted for the clinical parameters PEF, FEV₁, FVC and FEV₁/FVC.

The removal of the missing methods from the original set was based on no alteration of regression error upon their withdrawal, gross importance, measured by envelope-wise total split count, with 900 bagged trees, across all 9 clinical parameters, and weighting of the computational complexity and speed characteristics of the methods.

From the initial total of 54 envelopes and 486 extracted clinical parameters, the system suffered considerable shrinking until it produced only 8 envelopes and extracted 32 parameters (discarding the 5, least significant, types of measurements), without noticeable change in the regression results.

The final stage of the system consists of the classification of the regressed values, using a 5 fold cross-validated Random Forest model. After some experimentation, the best learner/classifier input was a combination of the 4 most relevant clinical parameters, as described earlier, and the patients' anthropometric parameters, height and age. The model, under these settings, attained around 8% and 10% misclassification rate, performing, respectively, as a two label and as a multiple label classifier. On the two label problem, the Random Forest was one among those with the highest precision and the one with the highest recall, among the best classifiers, with, respectively, 90.90% and 98.59%. Although the multiple label model could not accurately predict the

two restriction labeled observations, it managed to perform relatively well with Normal, Obstructive and Mixed abnormalities labels. The method chosen presented one of the best classification rates among all tested classifiers and was deemed the fastest and less computational intensive of the group.

In conclusion, a system design was devised and proved to be relatively precise in the classification of expiration recordings into both Normal and Abnormal labels and into Normal, Obstructive and Mixed labels. Some results of [13] and [6] were corroborated, such as the robustness of polynomial approximation of the expiration envelope. A study of several signal processing and machine learning methods was conducted and their characteristics were verified, such as, the usefulness of LPC in envelope calculation and the high classification performance of Random Forest.

There is significant room for improvement, starting with the collection of more recordings, with a higher repeatability quality. It would also be important to increase the number of recordings of restriction and mixed abnormalities labels, to produce more accurate regression and classification models. Also, further experiments could determine if other envelope generating methods could be removed without compromising both regression and classification results. Nevertheless, the methods that remain on the prototype are relatively simple and easy to implement on any smartphone app which makes the design feasible in practice. The results from this project seem promising and encourage further development.

5.2 Discussion

5.2.1 Main problems

As mentioned throughout the document, there were some project characteristic issues which are discussed in this section. First, some patient-wise issues are approached and then some technical issues are presented.

Patient's effort and fatigue: During the lung function screening, the patients have to perform several times the forced expiratory maneuver (FEM), or another forced maneuver. This can be due to the performance of the various tests or simply because of repeatability issues. Nevertheless, throughout this process, it is expected that a patient feels fatigue and vary their effort on the maneuvers. Consequently, either for spirometry as for the recordings, the expiration curves never quite exhibit the patient's real, undisturbed, expiration capacity, which leads to recording and spirometry effort mismatch.

Performing the FEM: One particularly hard part of the recording sessions is teaching patients how to correctly perform the FEM. In spirometry exams, the patient has to repeat the maneuver several times until at least three of them are considered acceptable according to the ATS standards. Additionally, expiring to a spirometer and to the screen of a smartphone have considerable differences. Due to the patients' breathing settling time, after a FEM, and the screening's tight schedule, for a considerable number of patients, only a couple of recordings were made per patient.

First, no mouthpiece was used on the smartphone tests, which forced the patients to, abnormally, keep their mouth and lips un-pursed, sometimes wide open, in order to unobstruct the passage of air from their mouths to the microphone. Second, patients have a tendency to either rotate the smartphone while performing the FEM or to flex their smartphone holding arm, which consequently changes the distance the air has to travel and alters the measured volume. Finally, it also seemed that exhaling without a mouth piece, made the patients uncomfortable enough to increase the probability of coughing, which is undesired for both spirometer and recording maneuvers.

Screening Schedule: Due to the recording session limitation to the ICAR screening sessions' schedule, only a relatively small number of patients was recorded. On the other hand, only relatively few different cases of abnormal lung function were recorded. This was, however, somewhat expected since a normal screening process is expected to find substantially more normal lung function patients than abnormal ones. Consequently, in the end only about 30% of the dataset contained abnormal observations, with obstruction lung disease's predominance over the other two abnormal labels.

Microphone saturation: The ICAR project consists of a nation wide study in which a team of physicians conducts their evaluations on an itinerant lab. As a consequence, some recordings were conducted outdoors, on somewhat windy conditions. On the other hand, patients have a tendency to rotate the smartphone, uncovering the microphone, while doing the FEM. Both situations lead to a set of recordings in which the microphone saturated, even though not always directly visible on the recordings.

Background noise during recordings: In addition to environmental noise, the recordings also would record the lab's background noise. This would include physicians speaking (even giving verbal incentive during the maneuvers), machine operation noises and occasional objects falling.

Recording application delay: Due to several unforeseen circumstances, the recording app design and implementation of all functionalities were only completely finalized mid-March. This made the recording sessions only possible from that point on. The closing of the dataset was done at about one month from the dissertation's deadline, in order to gather a reasonably comfortable number of recordings.

5.2.2 Future Work

For this dissertation, several methods, algorithms and experiments were implemented and conducted, in order to build a smartphone app that matched the requirements stated in Section 1.2. Nevertheless, some improvements and method optimizations can be done to further reduce the computing time and additional important features can be incremented. Some key features and other possibly relevant experiments are proposed.

5.2.2.1 Automatic Segmentation Algorithm Improvement

The segmentation algorithm worked relatively well, specially once the zero-time back-extrapolation was introduced. However, more recordings could have been used, from the original recordings set, if background and occasional noise could have been removed.

For this purpose, it would be of interest to experiment with Blind Source Separation algorithms in order to cancel such sound interferences.

5.2.2.2 AM demodulation

Another envelope generating method that could have been tested was AM demodulation. In terms of electronic circuitry the simplest implementation consists of a diode in series with a parallel of a capacitor and a resistance, which is not a particularly complex topology.

In terms of signal processing, it translates into taking the absolute values of the recordings and then filtering them with an LPF. Both concepts are already employed in the post-processing part (LPF) and on the Shannon envelopes, which take either the square or the absolute values and then calculate the logarithm of the curves.

Since it is a considerably accessible method and significantly lower complexity algorithm, the AM demodulation should also be experimented with.

5.2.2.3 Shape regression scheme

A relevant issue for spirometry that lacked in the project was the curve regression, mainly due to the inaccessibility to the spirometer's waveforms. In [19], one of the main focus was precisely the envelope regression stage since no classification was provided or even intended. The fact is, even if any application provided parameter regression and followed the standard ATS evaluation guidelines, when it came to the classification/diagnosis, the last word, concerning a patient's lung function would always be from the physician.

While this work attempted to provide a somewhat useful tool for lung function tracking, the physicians would still rather have all the clinical parameters, as well as the Flow-Volume and Volume-Time curves, to properly assess the patient, than relying solely of the parameters and a computerized diagnosis.

Due to these issues, the envelope regression would be a key feature in a future work. Incidentally, a slightly different architecture than of [19] could be useful. While they regress both Flow-Time and Volume-Time curves separately and then present both, it would be interesting if those two curves would be transformed in each other and calculate some sort of cross regression, which would then be presented to the user, to make up for the two distinct regression paths. Also, these final, cross-regressed curves could be used to correct the regressed clinical parameters, previously used to normalize the raw envelopes.

5.2.2.4 Robustness algorithms

One problem that affected a considerable amount of recordings was microphone saturation, as previously mentioned in this chapter. Although the recordings would not always present a significant spike, which would present maximum excursion values, there was often a wind-like sound followed by a subsequent reduction of sound magnitude. This effects caused a sharp decrease of the normal expiration sound volumes.

One of the most frequent causes of this effect was the exhalation incorrectly directed at the microphone instead of the screen.

Although, unverified, this problem might still be visible on the recordings as a temporary, DC offset. If this is accurate, then a detection algorithm for these events could be easily devised. This would be helpful since the application could then inform the user that the previous recording did not meet the standards and ask to perform and record the FEM again. The detecting algorithm could consist of calculating a moving average, over the recording itself, which if it crossed a certain absolute tolerance then it indicated the offset.

5.2.2.5 Bronchodilator change test

Considering a smartphone app implementation of this work, a bronchodilator (BD) test would be most significant for asthma and COPD patient's. A BD test may enable to monitor the lung change characteristics. It generally consists of performing two times the FEM, one before and another a few minutes after the inhalation of a BD. The differences between the two results should provide useful information about the state progression of the clinical parameters.

For instance, the non-improvement of FEV_1 values, may indicate a certain type of lung problem such as COPD. For asthma patients, the non-improvement of their PEF values (at the same time of day), may indicate that their prescription drug (BD) may not be the most suited treatment for that person.

Even if a patient has initial values consistent with the references (i.e. normal values), there still may be an obstruction which could be reduced with a bronchodilator. Due to the importance of this analysis, this is a basic step on physicians protocol towards lung function testing.

5.2.2.6 Forced Inspiration test

One of the first tasks in this project was to remove the inspiration sound and to keep only the exhalation sound. This was done for two main reasons: first the physicians feel that expiration maneuver provides more valuable information about the lung function than the inspiration, though they do both and may indicate different pathologies in case of an abnormality.

Second, it was assumed that inhalation produced a significantly quieter sound than expiration. In some cases, this was true, however, in the majority of the recordings the inspiration is louder than the expiration. Most inhalation sounds were found with a better clarity, less hesitation and left few doubts where the FEM portions started and ended, unlike expiration.

Additionally, people have a harder time doing the expiration part than when they do the inspiration portion of the FEM. In fact, when they expire, they have a higher tendency to cough than when they are inspiring.

Therefore, stands to reason to investigate the possibility of using the sounds of both directions of breathing since the inspiration also provides useful information about the lung function.

5.2.2.7 Comparison with predicted values

As was discussed on Section 4.6.4, height and age are important parameters for lung function. In fact, spirometry studies lead people to propose expected reference curves and spirometric values for healthy individuals [39]. Current physicians usually evaluate lung function based on FEV₁/FVC index, FEV₁-pre%, which stand for the ratio between the FEV₁'s predicted value and measured before the usage of bronchodilator, and similar types of ratios for FVC and PEF, either before and after the use of bronchodilator.

An important study that should be conducted would be to include such parameters into the feature space and observe the model's learning accuracy.

Appendix A

Initial Method Discussion

A.1 Choosing the Signal Processing methods

The initial literature review presented some relevant envelope generating methods, as well as pre and post processing algorithms to alter the sound signals and the envelopes. In this section such methods are discussed and the reasons for their use are presented.

The *Hilbert Transform* seemed to be a crucial process since it provides a simple way to compute the envelope of the breath signal. Another way to compute an envelope was tracking the major frequencies of the signal, with an *FFT*, and regenerating a time series. Also, the source power of the exhalation could be estimated using *Linear Predictive Coding*. Envelope signals might need to be filtered of noise in order to correctly extract the features. Therefore, it could prove beneficial to use *LPF*. All of these were employed by [19] and were found to give good results.

On the other hand, an envelope can be made immune to noise if a *function approximation* is computed. An initial choice among these functions had to be made in order to reduce the number of methods with the same approximate properties. The reasoning behind this choice was somewhat similar to the one [6] presented:

The *Linear Regression* probably would not produce good results since the FEM does not follow a linear function. A *Polynomial Regression* seems to be a good candidate though it has the instability problem for higher orders. On the other hand, it could have good results for an order between 3 and 5, which seems to be enough to represent the FEM accurately. *Splines* have a track record of lower order requirements and have similar results to Polynomial Regression. Though, since the polynomial order seems to be relatively low it might not be justifiable. In the end, Polynomial approximation was the one chosen.

The last couple of methods were envelope generating functions that shaped the envelope in a way that intensifies certain amplitudes. These are Shannon Energy and entropy. They have been used on heart rate signals and also had significant success [25, 26]. Additionally it would be interesting to compare them with the *Hilbert Transform* results.

It was defined that one of the requirements of the application is to have a processing time relatively small so only a reasonably few of the methods mentioned were expected to be present

on the final implementation. Primarily, all the signal processing methods were chosen because they presented different characteristics and their individual contribution was deemed worthwhile to experiment.

A.2 Choosing the Machine Learning methods

In this section a discussion concerning some machine learning algorithms is presented as well as the reasons why certain methods were chosen for this dissertations' experiments. Figure A.1 shows a comparison of several machine learning methods as presented on [3].

A.2.1 Regression

The regression portion of the initial system constituted a bottleneck in terms of learning speed and model implementation complexity. Furthermore, initially it was already expected that the dataset would not contain a substantial number of observations and, therefore, the quickness of learning in respect to the number of instances was also a pressing issue. The regression models considered were Regression Trees, Artificial Neural Networks (ANNs), Locally Weighed Regression and Support Vector Regression models (SVRs). Although Table A.1 presents a table with a comparison of classification models, some of them are also available for regression tasks with a minor alteration.

Regression Trees are fairly weak learning models that would most definitely overfit the data. However, if ensemble methods such as Bagging and Random Trees were used than that overfitting characteristic would benefit the whole tree set. These methods are relatively easy to understand, quick to train and use and are tolerant to missing values and irrelevant and redundant data.

ANNs, unlike tree methods, are strong learners that are quick give results and are generally quite accurate. However, the speed at which they learn is considerably low and they need a considerable learning dataset to become accurate, which would not be available.

Locally Weighed Regression is a k-Nearest Neighbors based regression model and, therefore, shares almost all it's characteristics. Although it is relatively transparent and learns instantaneously, the speed at which it classifies is proportional to the number of training examples and may be unaffordable in terms of memory usage, since it needs to keep every training example to compute the output.

Support Vector Regression models [40] attempt to minimize the generalized error bound in order to produce a generalized performance and avoid overfitting, unlike regular regression algorithms that minimize the observed training error. This is a promising method, however, it seemed a far more complex than needed for the application regression stage.

In light of of these options and characteristics, the Tree Bagging and Random Forest methods were chosen to regressed.

A.2.2 Classification

Choosing the methods to test on the classification portion of the project is a somewhat simpler task. Since by this stage the several clinical parameters are reduced to a much smaller set, several methods can be experimented upon, even if only for obtaining control values. The classification models considered were Classification Trees, Artificial Neural Networks, k-NN, Naïve Bayes and Support Vector Machines (SVMs). Some of these classifiers' reasoning can be found on the previous section since their classification counterparts have similar features.

Concerning Classification Trees, they present the same characteristics as regression trees, except for the fact that their output are labels instead of continuous values. In addition to the already mentioned ensemble methods, it seemed interesting to also use boosting, for which the AdaBoost algorithms were experimented with.

Naïve Bayes is a classification model which learns fairly quick, generates a moderate number of classifier parameters (mean and standard deviation for each class and feature). On the other hand, it generally has a relatively low performance and does not tolerate well irrelevant and redundant values. Additionally, most parameters of the feature space are expected to be highly interdependent, which further worsens the Naïve Bayes results' expectations. Nevertheless, it was also used as a control model from which to measure the other methods improvement.

As mentioned on the previous section, SVM classifiers are very accurate and quick to produce an output. Additionally, they tolerate very well irrelevant, redundant and interdependent features. Considering all of this and due to the more unconstrained situation, SVMs were used.

	Decision Trees	Neural Networks	Naïve Bayes	kNN	SVM	Rule-learners
Accuracy in general	**	***	*	**	****	**
Speed of learning with respect to number of attributes and the number of instances	***	*	****	****	*	**
Speed of classification	****	****	****	*	****	****
Tolerance to missing values	***	*	****	*	**	**
Tolerance to irrelevant attributes	***	*	**	**	****	**
Tolerance to redundant attributes	**	**	*	**	***	**
Tolerance to highly interdependent attributes (e.g. parity problems)	**	***	*	*	***	**
Dealing with discrete/binary/continuous attributes	****	***(not discrete)	***(not continuous)	***(not directly discrete)	** (not discrete)	***(not directly continuous)
Tolerance to noise	**	**	***	*	**	*
Dealing with danger of overfitting	**	*	***	***	**	***
Attempts for incremental learning	**	***	****	****	**	*
Explanation ability/transparency of knowledge/classifications	****	*	****	**	*	****

Table A.1: Comparing learning algorithms (**** stars represent the best and * star the worst in performance). Reproduction from [3].

References

- [1] Rob Pierce. Spirometry: an essential clinical measurement. *Australian family physician*, 34(7):535–9, July 2005. URL: <http://www.ncbi.nlm.nih.gov/pubmed/15999163>.
- [2] Gregg L. Ruppel. *Manual of Pulmonary Function Testing*. Mosby, 9th edition, 2008.
- [3] S. B. Kotsiantis. Supervised Machine Learning: A Review of Classification Techniques. In *Proceedings of the 2007 Conference on Emerging Artificial Intelligence Applications in Computer Engineering: Real World AI Systems with Applications in eHealth, HCI, Information Retrieval and Pervasive Technologies*, volume 31, pages 249–268, Amsterdam, The Netherlands, The Netherlands, 2007. IOS Press. URL: <http://dl.acm.org/citation.cfm?id=1566770.1566773>.
- [4] Cisco. Cisco Visual Networking Index : Global Mobile Data Traffic Forecast Update , 2013 – 2018. Technical report, Cisco, 2014.
- [5] AfricanTelecomsNews. Africa Mobile Factbook 2012. Technical report, Africa & Middle East Telecom-Week, 2012. URL: http://www.africantelecomsnews.com/resources/Factbook_Africa_14Q2_form.shtml.
- [6] Bas van Stein. A Mobile Smart Care platform Home spirometry by using the smartphone microphone. Master’s thesis, Leiden University, Leiden, The Netherlands, 2013.
- [7] Antonio Paes Cardoso, J.M. Reis Ferreira, and P. Rui Costa. *O Médico de Família e a Avaliação da Função Respiratória na DPOC*. 2006.
- [8] N. F. Macia. Pneumotachometers. In J. G. Webster, editor, *Encyclopedia of Medical Devices and Instrumentation, Vol. 5*, chapter Pneumotach, pages 367–379. Wiley, New York, 2nd edition, 2006.
- [9] C. P. Criée, S. Sorichter, H. J. Smith, P. Kardos, R. Merget, D. Heise, D. Berdel, D. Köhler, H. Magnussen, W. Marek, H. Mitfessel, K. Rasche, M. Rolke, H. Worth, and R. A. Jörres. Body plethysmography—its principles and clinical use. *Respiratory medicine*, 105(7):959–71, July 2011. URL: <http://www.ncbi.nlm.nih.gov/pubmed/21356587>, doi: [10.1016/j.rmed.2011.02.006](https://doi.org/10.1016/j.rmed.2011.02.006).
- [10] A. Hommersom, P.J.F. Lucas, M.V. Velikova, G. Dal, J. Bastos, J. Rodriguez, M. Germs, and H Schwietert. MoSHCA – My Mobile and Smart Health Care Assistant. In *Healthcom’13 : 15th international conference on E-health networking, applications & services, October 9-12, 2013*, pages 188–192, Lisbon, Portugal, 2013. URL: <http://hdl.handle.net/2066/122429>.

- [11] World Health Organization. Trade, foreign policy, diplomacy and health: E-Health. URL: <http://www.who.int/trade/glossary/story021/en/>.
- [12] M Kay, J Santos, and M Takane. mHealth: New horizons for health through mobile technologies. *World Health Organization*, 3, 2011. URL: http://www.who.int/entity/ehealth/mhealth_summit.pdf.
- [13] Wenyao Xu, Ming-chun Huang, Jason J. Liu, Fengbo Ren, Xinchen Shen, Xiao Liu, and Majid Sarrafzadeh. mCOPD. In *Proceedings of the 6th International Conference on Pervasive Technologies Related to Assistive Environments - PETRA '13*, PETRA '13, pages 1–8, New York, New York, USA, 2013. ACM Press. URL: <http://doi.acm.org/10.1145/2504335.2504383><http://dl.acm.org/citation.cfm?doid=2504335.2504383>, doi:10.1145/2504335.2504383.
- [14] Stefanie Auf der Mauer, Samantha Chan, Peter Chhour, Josh Homa, Vy Pham, and Karthik Pisupati. Mashavu Spirometer Project, 2013. URL: <https://decibel.ni.com/content/docs/DOC-5837>.
- [15] Andrew Brimer, Abigail Cohen, Olga Neyman, Braden Eliason, and Charles Wu. Low Cost Spirometer - DEBUT. Technical report, Washington University in St.Louis, St.Louis, 2012. URL: http://s3.amazonaws.com/challengepost/zip_files/production/3622/zip_files/Low-CostSpirometerReport-DEBUT.pdf?1338659900.
- [16] Maarten van der Heijden, Bas Lijnse, Peter J. F. Lucas, Yvonne F. Heijdra, and Tjard R. J. Schermer. Managing COPD exacerbations with telemedicine. *Artificial Intelligence in Medicine Lecture Notes in Computer Science*, 6747:169–178, 2011. URL: http://link.springer.com/chapter/10.1007/978-3-642-22218-4_21, doi:10.1007/978-3-642-22218-4_21.
- [17] Maarten van der Heijden, Peter J. F. Lucas, Bas Lijnse, Yvonne F. Heijdra, and Tjard R. J. Schermer. An autonomous mobile system for the management of COPD. *Journal of biomedical informatics*, 46(3):458–69, June 2013. URL: <http://www.ncbi.nlm.nih.gov/pubmed/23500485>, doi:10.1016/j.jbi.2013.03.003.
- [18] C. William Carspecken, Carlos Arteta, and Gari D. Clifford. TeleSpiro: A low-cost mobile spirometer for resource-limited settings. In *2013 IEEE Point-of-Care Healthcare Technologies (PHT)*, pages 144–147. IEEE, January 2013. URL: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6461305<http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6461305>, doi:10.1109/PHT.2013.6461305.
- [19] Eric C. Larson, Mayank Goel, Gaetano Boriello, Sonya Heltshe, Margaret Rosenfeld, and Shwetak N. Patel. SpiroSmart: Using a Microphone to Measure Lung Function on a Mobile Phone. In *14th ACM International Conference on Ubiquitous Computing*, page 10, Pittsburgh, Pennsylvania, USA, 2012.
- [20] M R Miller, J Hankinson, V Brusasco, F Burgos, R Casaburi, A Coates, R Crapo, P Enright, C P M van der Grinten, P Gustafsson, R Jensen, D C Johnson, N MacIntyre, R McKay, D Navajas, O F Pedersen, R Pellegrino, G Viegi, and J Wanger. Standardisation of spirometry. *The European respiratory journal*, 26(2):319–38, August 2005. URL: <http://www.ncbi.nlm.nih.gov/pubmed/16055882>, doi:10.1183/09031936.05.00034805.

- [21] James L. Flanagan, Jont B. Allen, and Mark A. Hasegawa-Johnson. *Speech Analysis Synthesis and Perception*. Springer-Verlag, Berlin, Heidelberg, 3rd edition, 2008.
- [22] David A. Winter. *Biomechanics and Motor Control of Human Movement*. Wiley, 4th edition, 2009.
- [23] K M Bushby, T Cole, J N Matthews, and J A Goodship. Centiles for adult head circumference. *Archives of Disease in Childhood*, 67(10):1286–1287, October 1992. URL: <http://adc.bmj.com/cgi/doi/10.1136/adc.67.10.1286>, doi:10.1136/adc.67.10.1286.
- [24] A. Veltcheva, P. Cavaco, and C. Guedes Soares. Comparison of methods for calculation of the wave envelope. *Ocean Engineering*, 30(7):937–948, May 2003. URL: <http://linkinghub.elsevier.com/retrieve/pii/S0029801802000690>, doi:10.1016/S0029-8018(02)00069-0.
- [25] Honghai Zhu and Jun Dong. An R-peak detection method based on peaks of Shannon energy envelope. *Biomedical Signal Processing and Control*, 8(5):466–474, September 2013. URL: <http://linkinghub.elsevier.com/retrieve/pii/S1746809413000025>, doi:10.1016/j.bspc.2013.01.001.
- [26] N Marques, Rute Almeida, Ana Paula Rocha, and M Coimbra. Exploring the Stationary Wavelet Transform Detail Coefficients for Detection and Identification of the S1 and S2 Heart Sounds. In *Computing in Cardiology Conference (CinC), 2013*, pages 891–894, Zaragoza, Spain, 2013. URL: http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6713521&punumber%3D6695807%26sortType%3Dasc_p_Sequence%26filter%3DAND%28p_IS_Number%3A6712387%29%26pageNumber%3D3%26rowsPerPage%3D100.
- [27] M.Sabarimalai Manikandan and K.P. Soman. A novel method for detecting R-peaks in electrocardiogram (ECG) signal. *Biomedical Signal Processing and Control*, 7(2):118–128, March 2012. URL: <http://linkinghub.elsevier.com/retrieve/pii/S1746809411000292><http://www.sciencedirect.com/science/article/pii/S1746809411000292>, doi:10.1016/j.bspc.2011.03.004.
- [28] S Choi and Z Jiang. Comparison of envelope extraction algorithms for cardiac sound signal segmentation. *Expert Systems with Applications*, 34(2):1056–1069, February 2008. URL: <http://linkinghub.elsevier.com/retrieve/pii/S0957417406003927>, doi:10.1016/j.eswa.2006.12.015.
- [29] H Liang, S Lukkarinen, and I Hartimo. Heart sound segmentation algorithm based on heart sound envelopgram. In *Computers in Cardiology 1997*, volume 24, pages 105–108. IEEE, 1997. URL: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=647841>, doi:10.1109/CIC.1997.647841.
- [30] Hisashi Wakita. Direct estimation of the vocal tract shape by inverse filtering of acoustic speech waveforms. *Audio and Electroacoustics, IEEE Transactions on*, 21(5):417–427, 1973. doi:10.1109/TAU.1973.1162506.
- [31] Abraham Savitzky and M. J. E. Golay. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Analytical Chemistry*, 36(8):1627–1639, July 1964. URL: <http://pubs.acs.org/doi/abs/10.1021/ac60214a047>, doi:10.1021/ac60214a047.

- [32] B.M. Shahshahani and D.A. Landgrebe. The effect of unlabeled samples in reducing the small sample size problem and mitigating the Hughes phenomenon. *IEEE Transactions on Geoscience and Remote Sensing*, 32(5):1087–1095, 1994. URL: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=312897>, doi:10.1109/36.312897.
- [33] V N Vapnik. An overview of statistical learning theory. *IEEE transactions on neural networks / a publication of the IEEE Neural Networks Council*, 10(5):988–99, January 1999. URL: <http://www.ncbi.nlm.nih.gov/pubmed/18252602>, doi:10.1109/72.788640.
- [34] Jonathan Milgram, Mohamed Cheriet, and Robert Sabourin. “One Against One” or “One Against All”: Which One is Better for Handwriting Recognition with SVMs? In Guy Lorette, editor, *Tenth International Workshop on Frontiers in Handwriting Recognition*, pages 1–6, La Baule (France), 2006. Suvisoft. URL: <http://hal.inria.fr/inria-00103955>.
- [35] J. Sunil Rao and William J.E. Potts. Visualizing Bagged Decision Trees. In *KDD*, pages 243–246, Providence, Rhode Island, USA, 1997. URL: <http://www.aaai.org/Papers/KDD/1997/KDD97-050.pdf>.
- [36] Leo Breiman. Random forests. *Machine learning*, 45(1):5—32, 2001. URL: <http://link.springer.com/article/10.1023/A:1010933404324>, doi:10.1023/A:1010933404324.
- [37] Robert E. Schapire. The strength of weak learnability. *Machine Learning*, 5(2):197–227, June 1990. URL: <http://link.springer.com/10.1007/BF00116037>, doi:10.1007/BF00116037.
- [38] Yoav Freund and Robert E. Schapire. Experiments with a new boosting algorithm. In Lorenza Saitta, editor, *ICML*, pages 148–156, Bari, Italy, 1996. Morgan Kaufmann. URL: <http://web.eecs.utk.edu/~parker/Courses/CS425-528-fall12/Handouts/AdaBoost.M1.pdf>.
- [39] J. Roca, F. Burgos, J. Sunyer, M. Saez, S. Chinn, J.M. Antó, R. Rodríguez-Roisin, Ph.H. Quanjer, D. Nowak, and P. Burney. Reference values for forced spirometry. *European Respiratory Journal*, 11(6):1354–1362, June 1998. URL: <http://erj.ersjournals.com/content/11/6/1354>, doi:10.1183/09031936.98.11061354.
- [40] Debasish Basak, Srimanta Pal, and DC Patranabis. Support vector regression. *Neural Information Processing: Letters and Reviews*, 11(10):203–224, 2007. URL: http://www.researchgate.net/publication/228537532_Support_vector_regression/file/60b7d51c2a0d86d3d9.pdf.