Development of Machine Learning Algorithms for Screening of Pulmonary Disease

by

Christian Infante

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Master of Engineering in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2017

Author .....
Department of Electrical Engineering and Computer Science
August 28, 2017

Certified by .....
Richard R. Fletcher
Research Scientist, D-Lab
Thesis Supervisor

Accepted by .....
Christopher Terman
Chair, Masters of Engineering Thesis Committee

Development of Machine Learning Algorithms for Screening of Pulmonary Disease

by

Christian Infante

Submitted to the Department of Electrical Engineering and Computer Science
on August 28, 2017, in partial fulfillment of the
requirements for the degree of
Master of Engineering in Electrical Engineering and Computer Science

Abstract

Pulmonary diseases are a leading cause of death worldwide. Much of their burden disproportionately affects the developing world. The MIT Mobile Technology Lab has developed a Mobile Kit which screens and diagnoses COPD and asthma. In this thesis, we analyze and further develop tools in this kit.

All of the data for this thesis were collected as part of a large medical study with our partner, the Chest Research Foundation (CRF), in Pune, India. The data consisted of 325 patients (135 healthy, 76 asthma, 46 COPD, 29 allergic rhinitis, and 39 other). Among the asthma and COPD patients, 67 had allergic rhinitis. All patients were examined using a mobile diagnostic kit designed at MIT consisting of a mobile stethoscope, peak flow meter, and questionnaire. All patients were also examained using the convential gold standard pulmonary function testing (PFT) lab. The performance of our Mobile Kit platform was previously analyzed and presented in a prior Master's thesis.

Building on our group's prior work, in this thesis we present three main contributions: 1) we have created a classifier for a new disease category, allergic rhinitis, which accounts for roughly half of all respiratory clinic patients; 2) we have explored and anlayzed the value of cough sounds as a diagnostic tools for pulmonary disease; and 3) we have analyzed data from a pulmonary function testing lab which were collected in parallel with our group's Mobile Diagnostic Kit, and have compared the performance.

In the first section of this thesis, we created a classifier for allergic rhinitis diagnosis, using the same multi-layer classification structure as was used in our group's prior

work. This integrated classifier demonstrated moderate performance with AUCs ranging from 0.87 to 0.90. As a second approach, a standalone classifier was also explored, which produced much better results, with an AUC of 0.96. Going forward, we plan to use an independent classifier as part of our diagnostics.

In the second part of this thesis, we explored the value of cough sounds for pulmonary diagnosis. Various classifiers were created for the screening and diagnosis of pulmonary disease through the analysis of cough sounds. We first created a classifier for the detection of Wet and Dry coughs (which can indicate overall pulmonary health), which had a high classification performance but limited diagnostic value. We then explored the diagnostic value of specific physical features of the cough sounds, including kurtosis, variance, zero crossing irregularity, and rate of decay. the utility of these features were then analyzed both in isolation and integrated with other Mobile Kit tools. It was discovered that these cough sound features do have value as a simple diagnostic tool to distinguish between asthma and COPD, as well as basic pulmonary health; however, it was found that cough sounds alone provide less value than other diagnostic tools for providing disease-specific diagnosis. When integrated with the Mobile Kit tools, cough sounds only improved performance on lung sounds; otherwise, coughs did not have any added benefit. Given the ease of data collection, we demonstrated that cough sounds can play a role in simple disease screening for use with community health workers.

For the third major part of this thesis, we did a thorough analysis of pulmonary function testing (PFT) data, which is the gold standard for pulmonary disease diagnosis. The PFT laboratory tools included spirometry, impulse oscillometry, body plethysmography, and lung gas diffusion (DLCO). We first explored a multi-layer classification structure. Using this structure, the PFT machines produced good results on each classification layer: Healthy vs. Unhealthy [AUC=0.90 (0.04)], Obstructive (Obs.) vs. Non-obstructive [AUC=0.95 (0.05)], Obs. AR vs. Obs. Non-AR [AUC=0.72 (0.10)], COPD + AR vs. Asthma + AR [AUC=0.95 (0.15)], COPD vs. Asthma [AUC=1.00 (0.04)], Non-Obs. AR vs. Non-Obs. Non-AR [AUC=0.92 (0.12)]. These results are only moderately better than the results yielded by our Mobile Diagnostic Kit: Healthy vs. Unhealthy [AUC=0.98 (0.02)], Obstructive (Obs.) vs. Non-obstructive [AUC=0.96 (0.04)], Obs. AR vs. Obs. Non-AR [AUC=0.90 (0.06)], COPD + AR vs. Asthma + AR [AUC=0.93 (0.09)], COPD vs. Asthma [AUC=1.00 (0.00)], Non-Obs. AR vs. Non-Obs. Non-AR [AUC=0.87 (0.12)]. Although these results are moderately good, the compounded error represents an unacceptable level of misclassification.

As an alternative to the multi-layer classification structure, we explored the use of individual classifiers for each disease, which yielded much better results. For the PFT

data, the individual classifiers produced the following results: asthma [AUC=0.96 (0.04)], COPD [AUC=0.99 (0.03)], and allergic rhinitis [AUC=0.74 (0.08)]. For the Mobile Kit data, the individual classifiers produced the following results: asthma [AUC=0.90 (0.05)], COPD [AUC=0.94 (0.05)], and allergic rhinitis [AUC=0.96 (0.03)].

In summary, building on our group's prior work, in this thesis we have expanded the capability of our Mobile Diagnostic Kit to include allergic rhinitis, as well as improved the diagnostic specificity to account for co-morbidities (asthma + AR, COPD + AR). Although our multi-layer classifier design has value in providing diagnostic insight and feedback to clinicians, we recommend that future versions of our Mobile Kit also include individual classifiers for specific disease categories (asthma, COPD, allergic rhinitis, asthma + AR, COPD + AR) in order to improve performance.

Thesis Supervisor: Richard R. Fletcher
Title: Research Scientist, D-Lab

Acknowledgments

First and foremost, I would like to thank my supervisor, Rich Fletcher, for his tireless commitment to the work on which this thesis is based. His vast and eclectic source of knowledge inspired and motivated much of this work.

I would also like to thank everyone at the Chest Research Foundation (CRF) in Pune, India, for being wonderful hosts and sharing with me most, if not everything, I know about pulmonary disease. I would especially like to thank Dr. Rahul Kodgule, for his insight and kindness; Yogesh Thorat, for his dedication and devotion; Dr. Sundeep Salvi, for his knowledge; and Dr. Shrikant Pawar, for his hard work in planning the next clinical study.

At MIT, I would like to thank the Tata Center for Development and Design, for funding and guiding my work. I would especially like to thank Dr. Rob Stoner and Dr. Chintan Vaishnav, for their teachings in the Tata Seminar; and Dr. Nevan Hanumara, for all his advice.

Finally, I would like to thank my my mother and father, without whose guidance, love, and support, I would not be.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Burden of Pulmonary Disease

1.1.1 Worldwide

Pulmonary diseases are a leading cause of death worldwide. Lower respiratory infections, chronic obstructive pulmonary disease (COPD), lung cancer, and tuberculosis accounted for 14% of deaths worldwide in 2012, and 11% of disability-adjusted-life-years (DALYs) lost worldwide in 2008 [11, 7]. Lower respiratory infections alone were the leading cause of DALYs lost (5.4%) [7].

An estimated 235 million people suffer from asthma, and 200 million people suffer from COPD [5, 6]. Asthma is characterized by airway inflammation whose symptoms include coughing, wheezing, breathlessness, and/or chest pain. Asthma patients usually also suffer from allergies [13].

An estimated 180,000 people die annually from asthma, with over 80% of these deaths occurring in low and lower-middle income countries [5, 9]. Most of these deaths occur in middle-aged patients and are considered preventable (either due to

inadequate or lack of care) [9]. The burden of asthma is expounded in developing countries, where patients have limited access to medications and proper care [9].

COPD is characterized by restricted air flow due to damaged air sacs (alveoli) and/or inflamed airways [8]. COPD is not exclusively a smoker's disease, as previously thought, even though smoking is its leading cause [13]. Other factors, such as poor air quality and poorly ventilated indoor biomass cooking, also cause COPD [10]. It causes irreversible lung damage, though the prevention of further deterioration is possible with appropriate treatment and lifestyle changes.

Most of the information known about COPD (its prevalence, morbidity, and mortality) is from data obtained in affluent countries; however, over 90% of COPD deaths occur in low- and middle-income countries [6]. COPD disproportionately affects these regions due to increased exposure to harmful environmental factors, such as air pollution and fumes from indoor biomass cooking.

1.1.2 India-specific

High Prevalence

In India, non-communicable diseases accounted for 53% of all deaths in 2005. Of these, 7% were due to chronic respiratory disease [42]. It is believed that around 20 million people suffer from COPD, and 30 million from asthma, in India [43]. Additionally, the country is suffering from a significant growing percentage of COPD mortality, believed to be one of the highest in the world [44].

Tobacco Use

Although the cigarette use in India is considerable, a significant number of Indian people use tobacco of other forms, such as hookah (pipes), bidi, chewing tobacco,

and chillum.)

It is estimated that 275 million people use tobacco in India, many under the age of 15. There is an increase in hookah use within Indian young adults, with prevalence reports within 5-14% [49]. Studies have shown that hookah smoking leads to an increased risk of lung disease; for example, lung cancer has been found to be six times more likely in hookah smokers than non-smokers [41].

Air Pollution

Outdoor

Air pollution is another concern. India has 13 of the 20 most polluted cities in the world [37]. Since 2010, there has been a 30% increase in acute respiratory infections cases. Many doctors blame this on poor air quality [36].

Indoor

Another aspect of Indian life that might indicate why pulmonary disease is so prevalent is the use of biomass cooking stoves, many of which are used indoors. In India, 70% of homes use biomass fuels for cooking in kitchens with poor ventilation [39]. The disparity between rural and urban settings is stark: 90% of rural homes use biomass stoves, whereas 32% of urban homes do [40]. In the rural settings, the negative effects of biomass stoves tend to affect women more [38].

Shortage of Doctors

Another reason for the high burden of pulmonary disease on India is a shortage of medical professionals, especially in rural areas. According to the WHO, there are only 0.725 physicians per 1000 citizens, a figure that has decreased since 1991 (when India had 1.225 physicians per 1,000 citizens. As reference, the United States has

2.554 physicians per 1,000 citizens [58].

Alternative Medicine

A study of pulmonary patients from Asian-Pacific countries (including India) quantified the burden of pulmonary disease, mainly looking at the utilization of healthcare and cost to the patient. It found that patients with rhinosinusitis (sinus infection), COPD, or asthma most frequently used a general practitioner (GP), while patients with allergic rhinitis (AR, an inflammation in the nose caused by allergens) tended to visit a pharmacist or a traditional medicine practitioner. The study also showed the use of drugs for these patients. When the patients were interviewed, 80% indicated that they had taken medication (usually inhalers, antibiotics, or antihistamines) within the previous four weeks. One interesting note is the popularity of traditional medicine among AR patients (23% of them visited a traditional medicine practitioner).

The direct (like medications) and indirect (like lost work hours) annual costs for an average patient of a pulmonary disease was found to be, on average, 637 USD. The biggest component of this was productivity loss. It should be noted that most of people used in this study had full-time jobs, most likely leading to the large annual cost due to work productivity loss [35].

1.2 Gold Standard for Pulmonary Disease Diagnosis

The gold standard for the diagnosis of many pulmonary diseases, as per the American Thoracic Society (ATS) and the European Respiratory Society (ERS) is Pulmonary Function Testing (PFT) [45]. these instruments include:

i) *Body plethysmograph*: a transparent box in which the patient sits and breathes in and out of a mouthpiece. The pressure within the box can change, allowing for a determination of lung volume.

ii) *Spirometry*: a recording of the amount and rate of air breathed in and out over a period of time.

iii) *Diffusing capacity of the lungs for carbon monoxide (DLCO)*: a measurement of the how much oxygen passes through the lungs into the blood.

iv) *Impulse Oscillometry*: a measurement of pulmonary resistance and reactance.

Many people in developing countries are unable to undergo PFT, either because it is too expensive, not available, or is administered by someone without adequate training. Therefore, many of the diagnoses in these regions rely on stethoscopes and/or clinical histories. To further aggravate the situation, many patients in these regions use traditional, non-Western medicine.

The burden of these diseases, coupled with the inadequacies of current diagnostic procedures in the developing world, creates a distinct need for high-quality, accurate, and affordable tools. There is increasing interest in mobile health technology, specifically through smart phone applications. For example, India has recently launched the Mobile Health Program, training community health worker to use a mobile health platform to improve maternal and child care in rural regions [4].

1.3 The Need for AI-Based Diagnostic Support

In the developing world, there is a scarcity of pulmonologists. This creates a dilemma: even if these areas had access to Pulmonary Function Testing (PFT), there are not sufficient trained professionals to accurately interpret the results. Additionally, many

patients in these regions seek help from practitioners of traditional (non-Western) medicine, which does not appropriately cover pulmonary disease.

Diagnostic tools based on an artificial intelligence foundation would address the above concerns. For one, the tool would be built from an analysis of data validated by the current gold standard for diagnosis. Also, utilizing these tools would require no knowledge of how they were created. These tools would therefore empower individuals without a medical background to accurately screen for pulmonary disease.

1.4 Previous Work in AI for Pulmonary Diagnostics

1.4.1 MIT Mobile Dianostic and Screening Kit

Work done by a former MIT student, Daniel Chamberliain, culminated in a Mobile Kit to screen for COPD and asthma, which provide the laragest burden of pulmonary disease worldwide. The Mobile Kit is hosted entirely on an Android phone. There are three main components to the Kit:

(i) *Questionnaire*: An electronic questionnaire was created with the guidance of a pulmonologist, using various other questionnaires as guidance. The questionnaire ultimately captured basic patient information (sex, weight, height, etc.), as well as "the onset, duration, and progress of breathlessness, coughing, chest pain, fever, and nasal symptoms".

(ii) *Peak Flow Meter*: An inexpensive tool commonly used by asthma patients to determine lung performance. A test involves taking a deep breath and blowing into the tube; the results are in liters per minute. Various trials of this tool are recorded by the Mobile Kit. The data are collected via augmented reality.

(iii) *Electronic Stethoscope*: A custom-made stethoscope was designed, using a

microphone and an auxiliary port. It allows lung and cough sounds to be recorded directly onto the Android device.

Previous work showed that the questionnaire and the peak flow meter provide the best diagnostic guidance for distinguishing between COPD and asthma.

1.4.2 Commercial Products

There are commercial products aimed at the developed world that allow users to monitor pulmonary health. One such product is Eko Device's DUO, an electronic stethoscope and EKG which continually monitors a user for any cardiac abnormalities [56]. Another product is StethoMe, which is an electronic stethoscope aimed for quick check-ups on children to determine whether or not a doctor's visit is required [57]

Many of the solutions for pulmonary disease screening in developing countries (and of any disease in general) is to check symptoms through the completion of a questionnaire. There are questionnaire-based solutions that focus on specific diseases, COPD and asthma being the most common. For COPD, sample solutions include the Lung Function Questionnaire and the COPD Population Screener questionnaire [52, 53]. These tools ask a couple questions related to smoking, age, and symptoms; however, patients determined to be at risk are advised to complete diagnostic exams (i.e. spirometry). Similarly, asthma has tools such as the Asthma Control Test, the Asthma Control Questionnaire, and the Asthma Therapy Assessment Questionnaire. Unlike the COPD tools described above, some of these asthma questionnaires are also very lengthy, making them tedious [54].

Another proposed solution, specific to COPD, is COPD-6, a mobile tool which mimics a spirometer. However, creating an entirely new piece of hardware as a solution proves costly (the COPD-6 is priced around 100-125 USD) [50, 51]. Financially,

this solution would not be viable in developing areas.

This leads to the desire to integrate a solution within pre-existing hardware. Given the rising interest of mobile health technology in developing countries, the natural progression is to host pulmonary disease screening software tools on a mobile device.

The current research is in using machine learning to create these tools. One Australia-based group, ResApp, is trying to diagnose disease through voluntary coughs [46]. However, all of their work so far relies on involuntary coughs, which are more cumbersome to acquire (since a microphone must be placed near a patient for hours). It is unclear whether it is proper to use their results on involuntary coughs and associate them to voluntary ones. Research has been done within the medical community to analyze the diagnostic value of involuntary coughs, but work still must be done to determine the value of voluntary coughs [55].

1.5 Scope of Thesis

This thesis continues from the work of Daniel Chamberlain, who developed and analyzed the first version of our group's Mobile Kit. There are three main contributions from my work:

1) The Mobile Kit has been further validated by comparing its results to Pulmonary Function Testing (PFT) data

2) Disease classification has been expanded to include allergic rhinitis (AR)

3) A new data source (cough sounds) has been analyzed

From Daniel Chamberlain's work, the Mobile Kit's classification scheme followed

Figure 3-4.

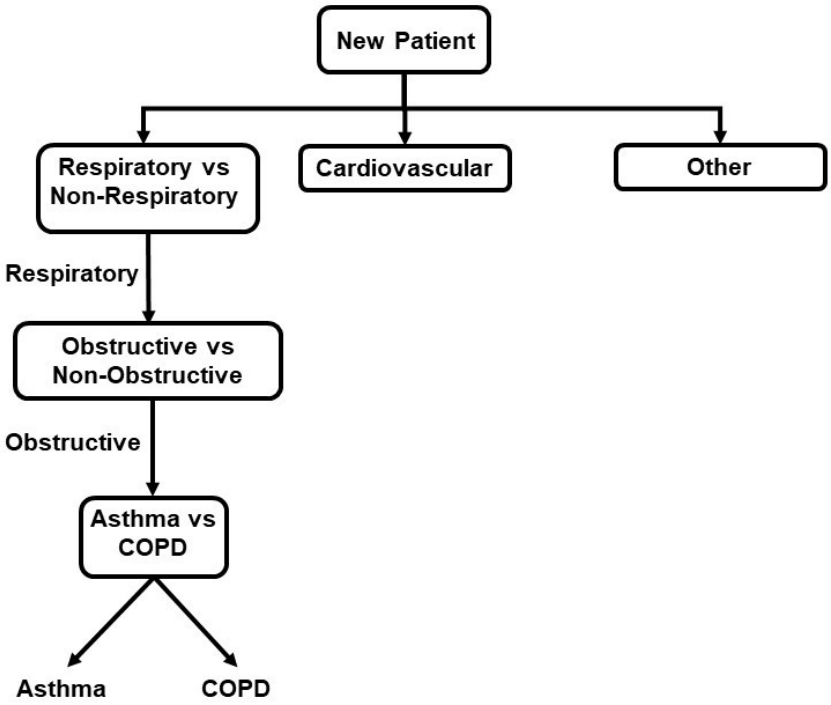

Figure 1-1: Previous classification scheme.

This classification scheme has been updated with the following:

1) A classifier to detect infective pulmonary diseases (tuberculosis, pneumonia, upper respiratory infection). While this thesis does not validate this classifier due to a lack of patients in the dataset with these diseases, the structure has been included for an upcoming study funded by the National Institutes of Health (NIH).

2) A classifier to differentiate between chronic (tuberculosis) and acute (pneumomnia, URI) infective pulmonary diseases. This classifier is not validated in this

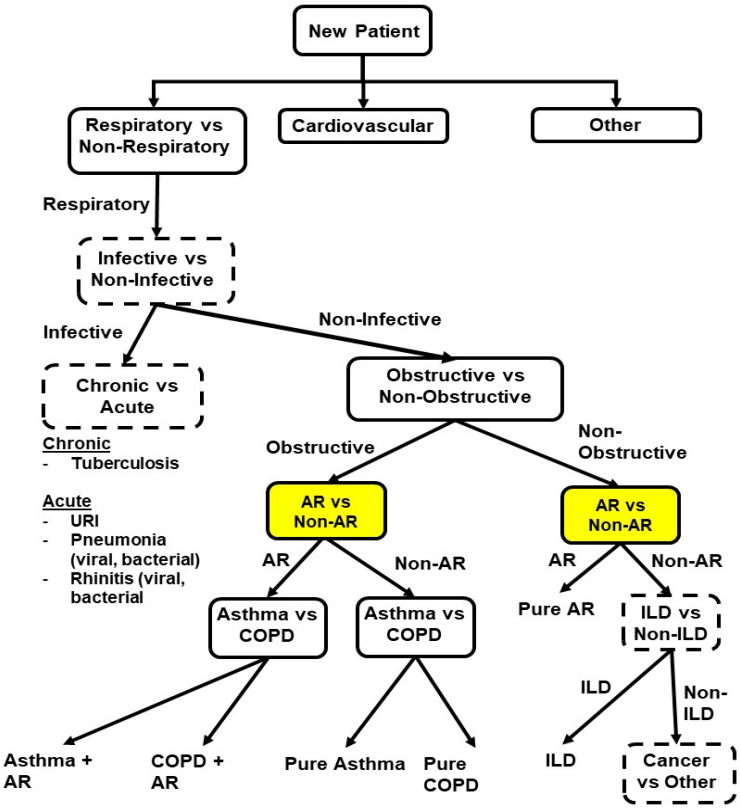

Figure 1-2: New classification scheme. Dotted lines denote classifiers that will be created in the future. Yellow boxes denote additions to the old scheme.

thesis, but the structure has been included for the NIH study.

3) Classifiers to detect interstitial lung disease (ILD) and cancer. These classifiers are not validated in this thesis, but the structure has been included for the NIH study.

4) A classifier to detect AR. Before, unhealthy patients were distinguished as having either an obstructive pulmonary disease (COPD or asthma) or a non-obstructive pulmonary disease (AR). Then, the patients classified as having an obstructive disease went through a COPD/asthma classifier. This design does not address the fact

that many COPD and asthma patients also have AR. Therefore, in the new classification scheme, all non-infective patients go through two classifiers: one that checks for AR, and another that checks for COPD/asthma.

This new scheme is summarized in Figure 4-4.

1.5.1 Overview

Chapter 2 summarizes the development of a new classifier which detects allergic rhinitis. First, an unsupervised analysis was done on the full risk factor and symptom questionnaire data to detect if an allergic rhinitis cluster emerged, and if so, what specific questions were useful to diagnose it. Second, a supervised analysis was done in which the new AR classifier was trained and validated. A coefficient analysis for this second portion determined the most important features (questions) from the questionnaire.

Chapter 3 analyzes cough sounds, a potential new feature source for the Mobile Kit. This analysis was done in three parts. First, an unsupervised analysis was done in search of any clusters within the data that corresponded to disease. Second, a supervised analysis was done for various classifiers, including Wet vs. Dry cough. Finally, the supervised analysis was repeated with the classifiers being trained on non-comorbid patients, but tested on both non-comorbid and comorbid patients. This was done under the hypothesis that non-comorbid patients would yield more accurate classifiers.

Chapter 4 analyzes raw Pulmonary Function Testing (PFT) data. We first summarize an unsupervised analysis. This was done in search of any inherent clusters, and if they emerged, how they correlated to disease. The second part of Chapter 2

analyzes the performance of data extracted from the PFT machines, both in isolation and in combination, when used to train our classifiers. These results were then compared to those of the Mobile Kit.

Chapter 5 summarizes all of the above findings, and suggests future improvements that can be made to the Mobile Kit implemented on a smart phone.

Chapter 2

Diagnosis of Allergic Rhinitis (AR)

Up to now, the Mobile Kit has focused on the diagnosis of COPD and asthma. However, allergic rhinitis is a common disease in developing nations, and oftentimes presents itself as a comorbidity. In this chapter, we address the diagnosis of allergic rhinitis. We begin by conducting an unsupervised analysis of the risk factor and symptom questionnaire (determined to be the most effective tool for the diagnosis of AR). We then integrate the AR classifer within the full diagnostic protocol. We conclude by analyzing an independent AR classifier.

2.1 Data Collection

Data were collected as part of a clinical trial hosted at a pulmonary research hospital (Chest Research Foundation) in Pune, India. Aside from the healthy controls, subjects were recruited from an equal sample of all patients arriving at the clinic exhibiting pulmonary symptoms. Table 3.1 shows the distribution of diseases within the dataset. The most common pulmonary diseases exhibited by patients were res-

piratory infections, Asthma, and COPD; this distribution is typical of that found in many developing countries.

All patients underwent our Mobile Kit's tools, which consisted of a clinical questionnaire, peak flow meter test, and auscultation (lung sounds) from 11 standard sites on the torso administered by a trained pulmonologist. The presence of any abnormal (adventitious) lung sounds at each site were noted manually by the pulmonologist.

Following the use of the mobile phone tools, each patient also underwent pulmonary function testing (PFT), which consisted of spirometry, body plethysmography, gas diffusion test (DLCO), and impulse oscillometry. Based on the PFT data and the clinical examination, the pulmonologist provided the final disease diagnosis, which was used as our labels for model training.

The protocol for this study received ethics approval from the appropriate boards at the Chest Research Foundation (Pune, India) and the Massachusetts Institute of Technology (Cambridge, USA).

Diagnosis	**Count**
No Pulmonary Disease	130
COPD Only	35
Asthma Only	47
Allergic Rhinitis Only	29
COPD + Allergic Rhinitis	11
Asthma + Allergic Rhinitis	56
Total	308

Table 2.1: Disease distribution within the cough analysis dataset

2.2 Feature Extraction

A description of each of the Mobile Kit's tools (developed and analyzed as part of Daniel Chamberlain's work), along with an explanation of the features extracted from each, is provided below. All of the extracted features were standardized to have zero-mean and unit-variance.

2.2.1 Risk Factor and Symptom Questionnaire

The Mobile Kit questionnaire was created with the aid of a pulmonologist to obtain details about the onset, duration, and progress of breathlessness, coughing, chest pain, fever, and nasal symptoms in a patient. The questionnaire also captures basic demographic information (age, sex, weight, etc.).

All binary responses were converted to Boolean variables, if they were not Boolean already. For example, the questionnaire asks if a patient is experiencing breathlessness– this question was converted to the feature Breathlessness Flag, True if the patient is breathless and False otherwise.

2.2.2 Peak Flow Meter

The second feature source consisted of five trials of a peak flow meter test. Each result can range from 0 to 800 L/min. Two features were extracted from these trials: 1) the maximum value, and 2) the ratio between the maximum value and the expected value given the patient's age, sex, and height (using equations 1 and 2 below). We call this feature "Peak Flow Meter Measure Over Reference".

(1) Male: -1.807 * Age (years) + 3.206 * Height (cm)

(2) Female: -1.454 *Age (years) + 2.368 * Height (cm)

2.2.3 Lung Sounds

During the 11-site auscultation, the pulmonologist noted the presence of abnormal (adventitious) lung sounds at each site (if any). The pulmonologist labeled the sound heard at each site as: Normal, Wheeze (polyphonic, monophonic), Crepitations (coarse, fine), Pleural Rub, or Squeak.

From these labels, 10 features were extracted. They are summarized in Table 2.2

Feature	**Description**
abnormal_sound_flag	True if at least 1 abnormal lung sound was noted at any of the 11 sites; False otherwise
num_abnormal_sounds	Total number of abnormal lung sounds heard across the 11 sites
wheeze_flag	True if at least 1 wheeze was noted at any of the 11 sites; False otherwise
num_wheezes	Total number of wheezes heard across the 11 sites
crackle_flag	True if at least 1 crackle was noted at any of the 11 sites; False otherwise
num_crackles	Total number of crackles heard across the 11 sites
num_wheezes_lower	Total number of wheezes heard in the lower lungs
num_wheezes_upper	Total number of wheezes heard in the upper lungs
num_crackles_lower	Total number of crackles heard in the lower lungs
num_crackles_upper	Total number of crackles heard in the upper lungs

Table 2.2: Summary of lung sound features

2.3 Classification Scheme

We updated the diagnostic protocol to account for allergic rhinitis. We did so by including an AR classifier among two subpopulations: obstructive and non-obstructive patients.

This scheme is summarized in Figure 2-1.

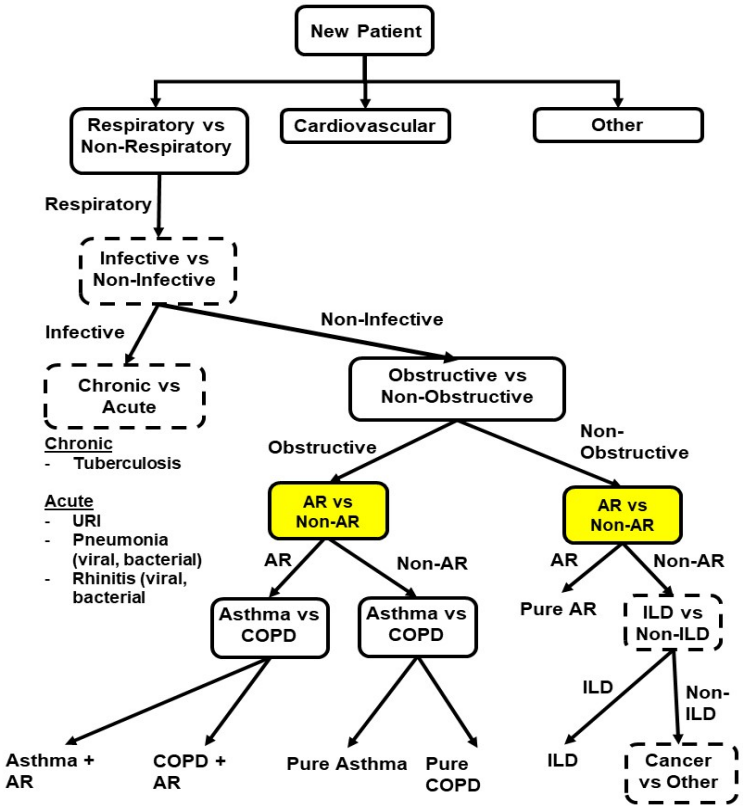

Figure 2-1: New classification scheme. Dotted lines denote classifiers that will be created in the future. Yellow boxes denote additions to the old scheme.

2.4 Ideal Tool for AR Diagnosis

In order to determine the best feature set for the diagnosis of allergic rhinitis, we trained two separate classifiers, one trained on all of the PFT data, and the other trained on all of the Mobile Kit tools. We did so under two scenarios:

1. We assumed the final classifier would be used within the full diagnostic protocol. Therefore, both classifiers had their testing and training sets extracted from patients with pulmonary disease.

2. We assumed the final classifier would be used independently. Therefore, both classifiers had their testing and training sets extracted from patients with and without pulmonary disease.

Tables 2.4 and 2.5 summarize the AR classifiers (under the obstructive and non-obstructive branches of the full diagnostic protocol) under Scenario #1. Table 2.3 summarizes the results of AR classifier under Scenario #2.

Features Used	**AUC**	**Sensitivity**	**Specificity**
PFT	0.69 - **0.74** - 0.77	0.50 - **0.71** - 0.86	0.56 - **0.71** - 0.86
Mobile Kit	0.94 - **0.96** - 0.97	0.94 - **0.94** - 1.00	0.87 - **0.89** - 0.92

Table 2.3: Performance of independent AR classifier trained on PFT vs. Mobile Kit

Features Used	**AUC**	**Sensitivity**	**Specificity**
PFT	0.65 - **0.69** - 0.73	0.50 - **0.80** - 0.90	0.45 - **0.64** - 0.82
Mobile Kit	0.87 - **0.90** - 0.93	0.85 - **0.92** - 1.00	0.73 - **0.80** - 0.87

Table 2.4: Performance of full diagnostic protocol's AR classifier (obstructive) trained on PFT vs. Mobile Kit

When the classifier is a part of the full diagnostic protocol (Scenario #1 above), utilizing either the Mobile Kit (median AUC: 0.89) or the PFT data (median AUC:

Features Used	AUC	Sensitivity	Specificity
PFT	0.88 - **0.92** - 1.00	0.75 - **1.00** - 1.00	0.67 - **1.00** - 1.00
Mobile Kit	0.81 - **0.87** - 0.93	0.67 - **0.83** - 1.00	0.67 - **0.89** - 0.89

Table 2.5: Performance of full diagnostic protocol's AR classifier (non-obstructive) trained on PFT vs. Mobile Kitnobs

0.93) result in similar performance under the non-obstructive branch (Table 2.5). However, for the classifier under the obstructive branch (Table 2.4, the Mobile Kit (median AUC: 0.90) has a clear advantage over PFT (median AUC: 0.69).

Additionally, when the classifier is standalone (Scenario #2 above), the Mobile Kit (median AUC: 0.96) has a clear advantage over PFT (median AUC: 0.74). From the supervised learning analysis of cough data, we determined that the questionnaire and the peak flow meter are the most useful tools for determining if a pulmonary disease is obstructive or not. Therefore, for the rest of this analysis, we focus on determining the efficacy of each questionnaire question for allergic rhinitis diagnosis.

2.5 Unsupervised Analysis

We began by running a factor analysis of the full questionnaire to determine the main features (or factors) it captures to determine how a future version of the tool could be shortened. We then ran a k-means cluster analysis on the full questionnaire in search of any clusters that correlated with allergic rhinitis, and if so, what risk factors and symptoms defined it.

2.5.1 Methods

Factor Analysis

Factor analysis assumes independent features. This condition failed, and so 5 features were removed (red sputum; colorless sputum; fever chills; fever rigors; fever sweating). This was due to a lack of positive examples in the dataset for these features. Additionally, the few positive examples tended to also be positive for some of the other features, creating a linear dependence.

The 65 features were pre-processed by their z-scores and the full feature space underwent the Kaiser-Meyer-Olkin (KMO) test for sampling adequacy, which attempts to determine if it is appropriate to explain the variance in dataset by a small number of factors. Generally, a value greater than 0.60 denotes appropriateness. Our dataset had a KMO score of 0.78.

In order to determine the appropriate number of factors, we used three methods: the Kaiser test, a Cattel scree plot, and an analysis of the cumulative variance. The Kaiser test states that all factors with eigenvalues greater than 1.00 should be kept. In our case, the test recommended we keep 17 factors. It should be noted that while this is a common technique, it has been criticized for overestimating the number of required factors.

We also used the Cattel Scree plot, which plots the number of features against their eigenvalues. Generally, these plots have a 'kink', after which point the eigenvalues level out. This method suggests the number of factors used is determined by this location. Figure 2-2 shows the scree plot for our data. It suggests using 10 factors. Like the Kaiser test, this method is criticized, specifically for being highly subjective.

Finally, we also analyzed the cumulative percent variance explained by the factors.

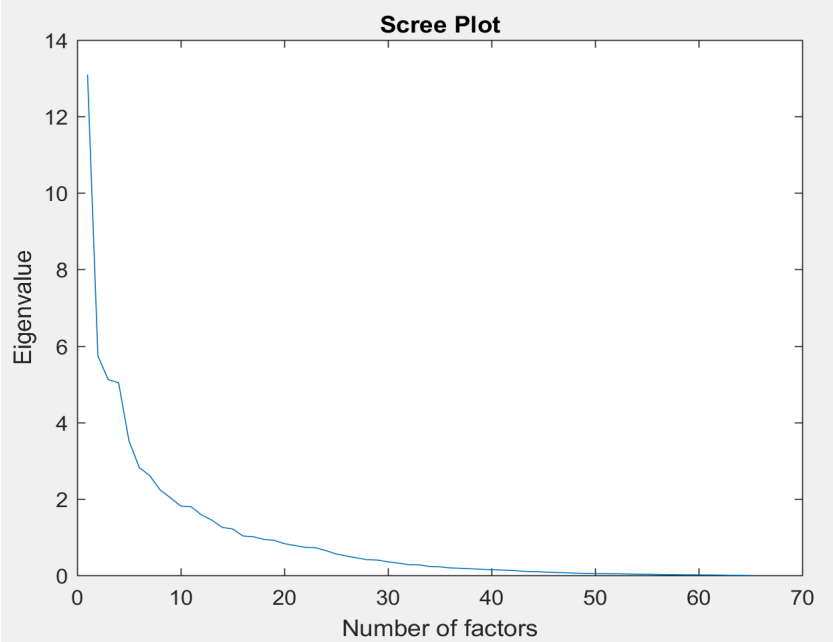

Figure 2-2: Cattel Scree plot from factor analysis of questionnaire data

The threshold for how much variance should be explained is research-dependent. Table 2.6 shows the minimum number of factors necessary to explain various amounts of data variance.

Percent of Variance Explained	50	60	70	80	90	95
Minimum Factors Required	5	8	11	16	24	31

Table 2.6: Results of variance test for determining number of factors for factor analysis of questionnaire data

Ultimately, we decided to run a factor analysis with 15 factors. Loadings less than 0.40 were suppressed.

k-Means Analysis

The data were standardized to have zero mean and unit variance. In order to determine the appropriate number of clusters, the mean silhouette score was calculated for cluster sizes ranging from 2 to 20 over 100 trials. The maximum silhouette score (0.21) occurred with 4 clusters.

2.5.2 Results

Factor Analysis

The following features failed to have a loading of at least 0.4 for any of the 15 factors: weight, height, paroxysmal nocturnal dyspnea, yellow sputum, green sputum, weight loss, loss of appetite, headache, somnambulance, family history of COPD, personal history of allergies, and alcohol intake. Of the 15 factors, 6 failed to have at least 3 loadings. The final 9 factors can be described as: Nasal Symptoms, Coughing, Breathlessness, Fever, Sputum Abnormalities, Chest Pain, Tobacco Use, Biomass Cooking, and Throat Symptoms.

K-means

The final k-means cluster analysis ran with 4 clusters. Table 2.7 summarizes the relevant aspects of each cluster. Cluster 1 was defined by a mean age of 62.49 years, and high incidences of chest pain (7.25%), family history of COPD (4.35%), and smoking (37.68%), relative to the other clusters. We initially believed this was indicative of a COPD cluster, given the older age of the group and the percentage of smokers. However, the group had almost even numbers of COPD patients (27.5%), asthma patients (34.7%), and healthy patients (34.7%). Yet, 54.3% of all COPD

patients in our dataset were included in this cluster. We ultimately believe that COPD did not arise as a dominant cluster due to a class imbalance; specifically, our dataset had many more patients with asthma and/or allergic rhinitis.

Cluster 2 was defined by low incidences of breathlessness (23.8%) and coughing symptoms (0.0%), relative to the other clusters. 83.5% of this cluster were subjects with no pulmonary disease, which we expected. Cluster 3 was small (5 patients). Interestingly, it captured all of the patients in the dataset with fever. We believe this is an outlier, potentially infectious, cluster. Cluster 4 was defined by high incidences of nasal symptoms (97.32%) and family history of allergies (41.96%). 78.5% of this cluster were patients with allergic rhinitis.

Feature	**Cluster 1**	**Cluster 2**	**Cluster 3**	**Cluster 4**
Age (yrs.)	62.49	41.11	44.20	50.71
Breathlessness (%)	92.8	23.8	80.0	89.3
Cough (%)	100.00	0.00	100.00	83.93
Chest pain (%)	7.25	3.28	40.00	3.57
Fever (%)	0.00	0.00	100.00	0.00
Nasal Symptoms (%)	0.00	2.46	80.00	97.32
Loss of appetite (%)	14.50	4.10	80.00	5.40
Family history of COPD (%)	4.35	0.82	0.00	1.79
Family history of asthma (%)	26.09	23.77	20.00	41.96
Smoke (%)	37.68	12.30	20.00	27.68
Number of Patients	69	122	5	112

Table 2.7: Summary of k-means cluster analysis on questionnaire features

Diagnosis	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Asthma + AR	1	0	2	53
COPD + AR	1	0	0	10
AR	0	4	1	24
COPD	19	6	0	10
Asthma	24	11	1	11
Healthy	24	101	1	4
Number of Patients	69	122	5	112

Table 2.8: Disesase composition of clusters from k-means analysis of risk factor and symptom questionnaire

2.5.3 Discussion

The k-means analysis shows clear clusters for subjects without pulmonary disease and patients with allergic rhinitis. A clear COPD cluster potentially did not emerge due to a lack of COPD samples in the data. Future iterations of this analysis should aim to acquire more data to balance the classes out. A clear asthma cluster also did not emerge. Among asthma patients, most appeared in cluster 1 (51%), and the rest appeared evenly between clusters 2 and 3. The confusion between asthma and allergic rhinitis is expected, given they oftentimes go hand-in-hand. Overall, this analysis predicted some pertinent features for diagnosing allergic rhinitis, specifically: age, smoking, family history of allergies, nasal symptoms, coughing, breathlessness, and fever.

Factor analysis provided evidence that it is possible to explain the dataset with 9 latent factors. We predict that a relatively short questionnaire which captures these main ideas should perform similar to the full questionnaire.

2.6 Independent Classifier

2.6.1 Methods

We retrained the independent AR classifier with only the questionnaire to confirm that it was the most useful tool in the Kit. We then performed a coefficient analysis to determine the most useful questions.

2.6.2 Results

Table 2.9 summarizes the results of the AR classifier trained solely on the questionnaire data. Figure 2-3 summarizes the classifier's top coefficients.

Features Used	AUC	Sensitivity	Specificity
Questionnaire	0.94 - **0.96** - 0.97	0.95 - **1.00** - 1.00	0.85 - **0.88** - 0.90

Table 2.9: Performance of independent AR classifier trained on the questionnaire

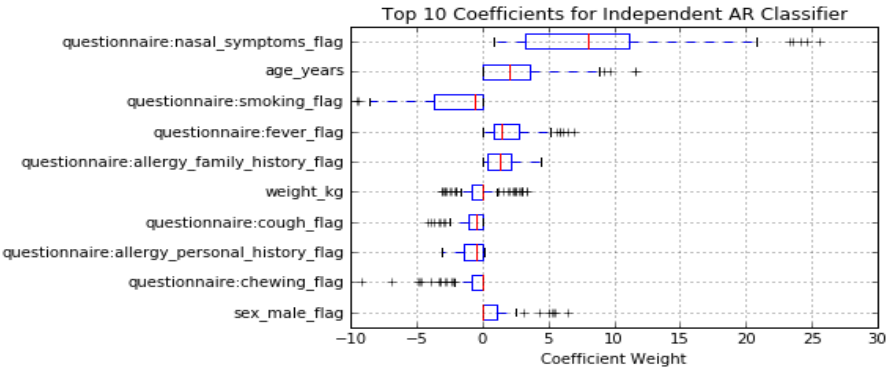

Figure 2-3: Top 10 coefficients of independent AR classifier when utilizing only the questionnaire

2.6.3 Discussion

As shown in Table 2.9, the independent classifier had comparable performance when it was trained with all of the Mobile Kit tools to when it was trained with just the questionnaire data. This further reinforces our previous finding that the questionnaire is the most useful feature for AR diagnosis, and an AR-specific questionnaire would be the ideal tool within the Mobile Kit.

The coefficient analysis (Figure 2-3) aligns with the findings from the unsupervised analysis. The most important features were determined to be Nasal Symptoms, Age, Smoking, Fever, and Family History of Allergy. This suggests that a very short questionnaire which captures these data may accurately diagnose allergic rhinitis.

2.7 Integration of AR Classification into Full Diagnostic Protocol

2.7.1 Methods

We restructured the old classification scheme by introducing allergic rhinitis classifiers. These are found under the Obstructive vs. Non-obstructive Pulmonary Disease branch; one handles patients labeled as having obstructive disease, while the other handles patients labeled as having a non-obstructive disease.

2.7.2 Results

Table 2.15 shows the results of the AR classifier under the obstructive branch. Table 2.13 shows the results of the AR classifier under the non-obstructive branch. For completion, we have updated the performance of the Mobile Kit by retraining and

testing the other classifiers in the new diagnostic protocol. They are summarized in Tables 2.10, 2.11, 2.12, and 2.14.

Features Set(s)	AUC	Sensitivity	Specificity
L	0.61 - **0.63** - 0.66	0.22 - **0.27** - 0.31	1.00 - **1.00** - 1.00
Q	0.97 - **0.98** - 0.99	0.91 - **0.93** - 0.98	0.94 - **1.00** - 1.00
P	0.83 - **0.86** - 0.88	0.55 - **0.70** - 0.80	0.81 - **0.94** - 1.00
Q + P	0.97 - **0.98** - 0.99	0.90 - **0.92** - 0.95	1.00 - **1.00** - 1.00
Q + L	0.98 - **0.99** - 0.99	0.91 - **0.95** - 0.98	1.00 - **1.00** - 1.00
P + L	0.85 - **0.88** - 0.90	0.60 - **0.68** - 0.78	0.88 - **0.94** - 1.00
Q + P + L	0.97 - **0.98** - 0.99	0.90 - **0.92** - 0.95	1.00 - **1.00** - 1.00

Table 2.10: Performance of Healthy vs. Unhealthy classifier utilizing different Mobile Kit feature set combinations under the new classification scheme

2.7.3 Discussion

For the AR classifier handling patients with obstructive disease (Table 2.15), the questionnaire (median AUC: 0.91) greatly outperforms lung sounds (median AUC: 0.59) and the peak flow meter (median AUC: 0.68). While the questionnaire achieves the best specificity of any combination of feature sets, it is not ideal (median specificity: 0.82) and has a large interquartile range (17%).

For the AR classifier handling patients with non-obstructive disease (Table 2.13), both the questionnaire and peak flow meter have comparable performance (for both, median AUC: 0.91). However, the questionnaire outperforms the peak flow meter in sensitivity (median sensitivity of 0.86, compared to 0.83 for the peak flow meter) and specificity (median specificity of 0.90, compared to 0.89 for the peak flow meter).

Features Set(s)	AUC	Sensitivity	Specificity
L	0.60 - **0.65** - 0.68	0.23 - **0.26** - 0.39	0.83 - **1.00** - 1.00
Q	0.87 - **0.89** - 0.92	0.60 - **0.77** - 0.90	0.83 - **0.83** - 1.00
P	0.88 - **0.91** - 0.95	0.71 - **0.79** - 0.89	0.80 - **1.00** - 1.00
Q + P	0.92 - **0.96** - 0.98	0.79 - **0.86** - 0.96	0.80 - **1.00** - 1.00
Q + L	0.90 - **0.94** - 0.96	0.68 - **0.81** - 0.90	0.83 - **1.00** - 1.00
P + L	0.88 - **0.93** - 0.96	0.71 - **0.82** - 0.89	0.80 - **1.00** - 1.00
Q + P + L	0.94 - **0.96** - 0.98	0.81 - **0.89** - 0.93	0.95 - **1.00** - 1.00

Table 2.11: Performance of Obstructive vs. Non-obstructive classifier utilizing different Mobile Kit feature set combinations under the new classification scheme

It is clear from the performance of these two classifiers that the questionnaire is the best Mobile Kit tool for detecting allergic rhinitis. However, a lower-than-ideal specificity, coupled with any error carrying over from higher branches, will decrease performace of the new diagnostic protocol when used in the field. Therefore, it is recommended that an independent classifier be used for the detection of allergic rhinitis, based on relevant questionnaire questions. The restructuring of the full diagnostic protocol to one of multiple, independent, and disease-specific classifiers is explored in the PFT analysis chapter.

Features Set(s)	AUC	Sensitivity	Specificity
L	0.56 - **0.63** - 0.66	0.83 - **0.92** - 1.00	0.25 - **0.38** - 0.63
Q	0.96 - **1.00** - 1.00	1.00 - **1.00** - 1.00	0.86 - **1.00** - 1.00
P	0.73 - **0.81** - 0.87	0.60 - **0.80** - 1.00	0.68 - **0.71** - 0.86
Q + P	0.98 - **1.00** - 1.00	1.00 - **1.00** - 1.00	0.96 - **1.00** - 1.00
Q + L	0.95 - **1.00** - 1.00	1.00 - **1.00** - 1.00	0.86 - **1.00** - 1.00
P + L	0.76 - **0.83** - 0.89	0.65 - **0.90** - 1.00	0.57 - **0.71** - 0.86
Q + P + L	1.00 - **1.00** - 1.00	1.00 - **1.00** - 1.00	1.00 - **1.00** - 1.00

Table 2.12: Performance of Pure COPD vs. Pure Asthma classifier utilizing different Mobile Kit feature set combinations under the new classification scheme

Features Set(s)	AUC	Sensitivity	Specificity
L	0.58 - **0.60** - 0.65	0.29 - **0.43** - 0.86	0.33 - **0.90** - 1.00
Q	0.86 - **0.91** - 0.95	0.71 - **0.86** - 1.00	0.79 - **0.90** - 1.00
P	0.86 - **0.91** - 0.96	0.79 - **0.83** - 1.00	0.78 - **0.89** - 0.89
Q + P	0.83 - **0.88** - 0.93	0.67 - **0.83** - 1.00	0.67 - **0.78** - 0.89
Q + L	0.84 - **0.90** - 0.94	0.71 - **0.86** - 0.86	0.79 - **0.89** - 0.90
P + L	0.85 - **0.89** - 0.94	0.67 - **0.83** - 1.00	0.78 - **0.89** - 0.89
Q + P + L	0.81 - **0.87** - 0.93	0.67 - **0.83** - 1.0	0.67 - **0.89** - 0.89

Table 2.13: Performance of AR classifier (non-obstructive) utilizing different Mobile Kit feature set combinations under the new classification scheme

Features Set(s)	AUC	Sensitivity	Specificity
L	0.67 - **0.70** - 0.77	0.33 - **0.58** - 0.67	0.73 - **1.00** - 1.00
Q	0.91 - **0.93** - 0.97	1.00 - **1.00** - 1.00	0.73 - **0.82** - 0.91
P	0.77 - **0.82** - 0.91	0.50 - **1.00** - 1.00	0.55 - **0.73** - 1.00
Q + P	0.90 - **0.94** - 0.98	**1.00** - 1.00 - 1.00	**0.73** - 0.82 - 0.91
Q + L	0.89 - **0.93** - 0.97	**1.00** - 1.00 - 1.00	**0.73** - 0.82 - 0.91
P + L	0.64 - **0.67** - 0.70	0.54 - **0.77** - 0.92	**0.40** - 0.53 - 0.75
Q + P + L	0.86 - **0.93** - 0.95	1.00 - **1.00** - 1.00	**0.72** - 0.82 - 0.91

Table 2.14: Performance of COPD vs. Asthma (comorbid) classifier utilizing different Mobile Kit feature set combinations under the new classification scheme

Features Set(s)	AUC	Sensitivity	Specificity
L	0.55 - **0.59** - 0.64	0.21 - **0.39** - 1.00	0.24 - **0.82** - 0.96
Q	0.87 - **0.91** - 0.94	0.86 - **0.93** - 1.00	0.71 - **0.82** - 0.88
P	0.65 - **0.68** - 0.72	0.62 - **0.77** - 0.92	0.40 - **0.53** - 0.68
Q + P	0.89 - **0.92** - 0.95	0.85 - **0.92** - 1.00	0.73 - **0.80** - 0.87
Q + L	0.87 - **0.90** - 0.93	0.86 - **0.93** - 1.00	0.71 - **0.81** - 0.88
P + L	0.64 - **0.67** - 0.70	0.54 - **0.77** - 0.92	0.40 - **0.53** - 0.75
Q + P + L	0.87 - **0.90** - 0.93	0.85 - **0.92** - 1.00	0.73 - **0.80** - 0.87

Table 2.15: Performance of AR classifier (obstructive) utilizing different Mobile Kit feature set combinations under the new classification scheme

Chapter 3

Use of Cough Sounds for Screening and Diagnosis of Pulmonary Disease

In this chapter, we analyze the screening and diagnostic value of adding cough sound analysis to the Mobile Kit. We do so in four parts. First, we summarize an exploratory analysis involving unsupervised learning to detect any inherent clusters that map cough sound features to disease. Second, we analyze the value of coughs in isolation via a supervised learning approach, first to discern between wet and dry coughs, and then to discern between various levels of pulmonary disease screening and diagnosis. Third, we integrate cough sounds with the other Mobile Kit tools (questionnaire, peak flow meter, and lung sounds) to analyze the use of coughs as an additional feature. Finally, we extend the supervised learning analysis by training the classifiers with non-comorbid patients, and seeing if there are any improvements when compared to the classifier trained on both comorbid and non-comorbid patients.

3.1 Motivation

Cough is one of the earliest symptoms of many pulmonary diseases. It is also very simple to record, not requiring any special tools. Whereas the recording of lung sounds in the current Mobile Kit requires the use of an electronic stethoscope, coughs can be recorded directly with the phone. This can be of use in resource-lacking areas where the use of our current kit's peripherals (like the peak flow meter or the stethoscope) might be too costly, or too difficult (for example, a very young child might have trouble completing the peak flow meter test).

A cough is generally categorized as wet or dry. However, these categories are highly subjective, with wet coughs tending to be characterized by the presence of phlegm, and dry coughs characterized by the lack of phlegm. Given the subjectivity of classification, it is very enticing for pulmonologists to have a tool that can accurately (and automatically) classify coughs.

Knowing whether a cough is wet or dry can give some insight into a patient's pulmonary health. For example, a wet cough can indicate conditions like lower respiratory tract infections, pneumonia, and bronchitis, while dry coughs can indicate conditions like allergies and asthma. Recent work on cough analysis has primarily focused on naturally occurring involuntary coughs. This form of data collection requires continuous recording, often over many hours, to ensure a sufficient number of coughs are collected [33, 34]. However, for the purpose of a screening tool, it is more relevant to study voluntary coughs that can be readily acquired. Many groups who are trying to create a diagnostic tool using voluntary coughs are extending on their previous work on involuntary coughs.

3.1.1 Previous Work

Aside from wet/dry classification, there are three main areas of research within the cough community:

Segmentation

Cough segmentation involves the automatic extraction of individual coughs from a recording containing multiple coughs. Many studies involving cough sounds manually segment coughs. However, for use in final products, automatic segmentation methods are required. Much of the recent work in cough segmentation involves complex, computationally-intensive models, such as neural networks.

Detection

Cough detection involves the automatic classification of a sound as a cough. This is of great interest to the smart home community, especially with the growing work being done in telemedicine. The general application within this context is to continuously record sound and be able to detect when a user coughs. Just like cough segmentation, many of the recent approaches are computationally expensive. However, this process is not a requirement for our Mobile Kit.

Analysis

Cough analysis involves a combination of signal analysis and machine learning. Much of the work done in this area involves: a) extracting relevant time- and frequency-domain features from cough signals, and b) creating machine learning classifiers to

screen for various pulmonary disease (asthma, COPD, allergic rhinitis, pneumonia, etc.).

3.2 Data Collection

Cough recordings (30 seconds in length) were captured from each patient at the trachea, during which time patients were asked to cough multiple times. Each individual cough used for training was also labeled as Wet (containing phlegm) or Dry (not containing phlegm). Since cough sound data were only available for a portion of the patients in our study, we used a sample of a larger data set which was used for our larger diagnostic study. The disease distribution within the dataset is shown in Table 3.1. Summary statistics are shown in Table 3.2.

Diagnosis	**Count**
No Pulmonary Disease	33
COPD Only	7
Asthma Only	15
Allergic Rhinitis Only	11
COPD + Allergic Rhinitis	4
Asthma + Allergic Rhinitis	17
Total	87

Table 3.1: Disease distribution within the cough analysis dataset

Statistic	Value	Statistic	Value
Male (%)	52.05	Family History of COPD (%)	0.00
Age (years)	46.34	Family History of Allergies (%)	32.88
Weight (kg)	61.38	Personal History of Allergies (%)	15.07
Breathless (%)	58.90	Exposed to Biomass Cooking (%)	13.70
Coughing (%)	42.47	Smoke (%)	19.18
Chest Pain (%)	42.47	Chew Tobacco (%)	26.03
Fever (%)	1.37	Consume Alcohol (%)	9.59
Nasal Symptoms (%)	35.62	Max Peak Flow Meter Reading (L/min)	296.71

Table 3.2: Summary statistics of the cough analysis dataset

3.3 Data Preprocessing

From each cough sound recording, the first complete cough (defined as having 100 milliseconds of silence before and after the cough event) was extracted manually using the Audacity software. For some patients, it was impossible to detect a single, distinct cough (usually due to uncontrollable coughing episodes common in COPD or asthma patients). These patients were discarded from the analysis.

Figure 3-1 shows a sample signal of a full recording lasting 30 seconds, containing seven complete coughs; Figure 3-3 shows the extracted cough signal.

3.4 Cough Feature Selection

3.4.1 Segmentation

The first step in cough analysis is to find the presence of a cough within the recorded sound file. We implemented a cough segmentation algorithm which extracts the first complete cough in a recording using the following algorithm:

1. The original sound file (Figure 3-1) is smoothed by applying local regression (weighted least squares with a 2nd degree polynomial model) using a span of 2% of the data. The signal is then normalized to span from 0 to 1. Figure 3-1 shows an example of a complete recording. Figure 3-2 shows the smoothed signal.

2. A peak detection algorithm is applied to the smoothed signal (Figure 3-2) to find all cough peaks present in the file.

3. In order to select the first complete cough in the series, each peak is analyzed individually. The zero-crossing of the first derivative is used to determine the starting point of the cough, and the slope of the trailing edge of the cough is used to determine the end of the cough.

4. The flatness of the slope is used as a criterion to determine if the cough segment is complete. If the algorithm encounters a new peak before the previous cough sound has settled (defined as the slope being below a pre-determined threshold), then the algorithm discards the current cough segment and starts a new search using the next available cough peak.

3.4.2 Feature Extraction

Once the leading cough segment was extracted from a recording, the next step is to analyze the sound and extract specific features. Figure 3-3 shows the automatically

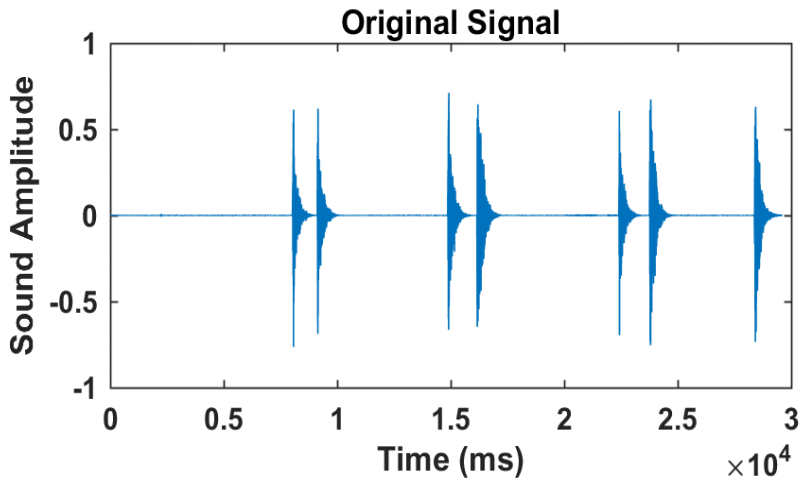

Figure 3-1: Sample sound file showing raw cough data

extracted cough signal from the original signal (Figure 3-1).

Starting from an initial set of approximately 30 features published in the literature, we selected features which have been previously used for diagnostic prediction and also gave good results on initial trials. They are summarized in Table 3.3.

3.5 Classification

3.5.1 Wet vs. Dry Cough

The first classification we performed with the cough data was to discern between wet (containing phlegm) and dry (not containing phlegm) coughs.

3.5.2 Pulmonary Health and Disease

Given the fact that much of the data had to be removed due to the inability to extract a single valid cough from patients' cough recordings, we followed the classification

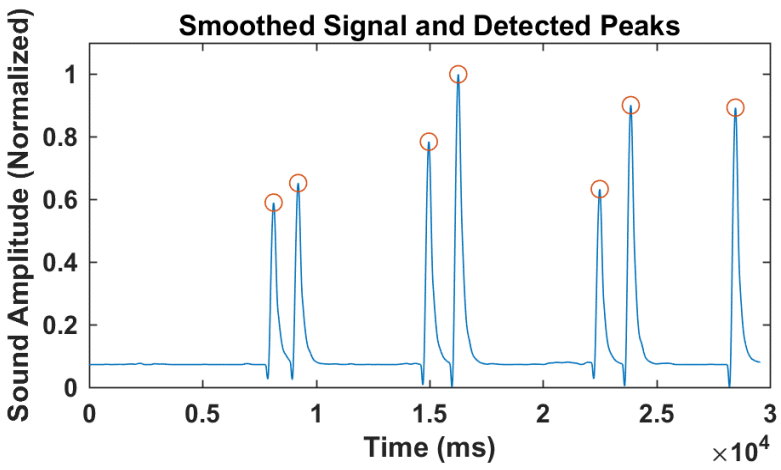

Figure 3-2: Smoothed cough signal magnitude and detected peaks (circled)

scheme used in Daniel Chamberlain's work. It is a tree-like structure composed of three layers. The first layer determines whether a patient has a pulmonary disease or not (Respiratory vs. Non-Respiratory, or Unhealthy vs. Healthy). Respiratory patients then go through the next layer of classification which determines whether they have an Obstructive or Non-obstructive pulmonary disease. Obstructive patients then go through the final layer, which determines whether they have COPD or asthma. For this analysis, at any given layer, we assume perfect classification from the previous layer, if applicable. For example, we assume that the Obstructive classifier is perfect, and only train the COPD vs. Asthma classifier on patients known to have Obstructive pulmonary disease.

This scheme is summarized in Figure 3-4.

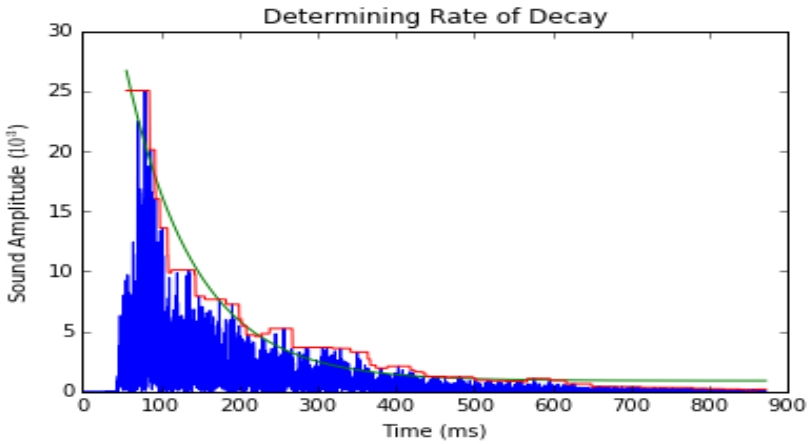

Figure 3-3: Plot of extract cough segment (blue); computed upper envelope (red); exponential fit (green)

3.6 Unsupervised Analysis

3.6.1 Methods

In order to explore any potential hidden correlations between cough features and disease, we performed a standard k-means clustering analysis. The ideal number of clusters was determined to be three (by averaging the silhouette score over a range of cluster sizes from two to ten, over 100 trials).

3.6.2 Results

The results of the cluster analysis are shown in Table 3.4. Table 3.5 show the average feature values (normalized) within each cluster.

Feature	Description
Kurtosis	The fourth-order moment of the signal, computed from the magnitude, |x(t)|, which is a measure of its "Gaussianity". This feature has been used to automatically detect pertussis (whooping cough). [24]
Variance	The variance of the signal's magnitude. This feature has been used to detect abnormal pulmonary function. [20]
Zero cross irregularity	A measure of the deviation between time intervals in which the cough signal crossed the x-axis. This measure has been used in previous analyses to detect wheezes. [15]
Rate of decay	The exponent value of an exponential curve fitted to the magnitude of the cough signal, |x(t)|: Figure 3-3 shows the computed upper envelope and the fitted curve for the automatically extracted cough described above.

Table 3.3: Description of extracted cough sound features

Diagnosis	Cluster 1	Cluster 2	Cluster 3
Asthma	7	0	5
Asthma + Allergic Rhinitis	4	4	7
Allergic Rhinitis	4	3	4
COPD	1	1	4
COPD + Allergic Rhinitis	0	2	2
Healthy	13	1	17
Total	29	11	39

Table 3.4: Results of k-means cluster analysis utilizing cough sound features

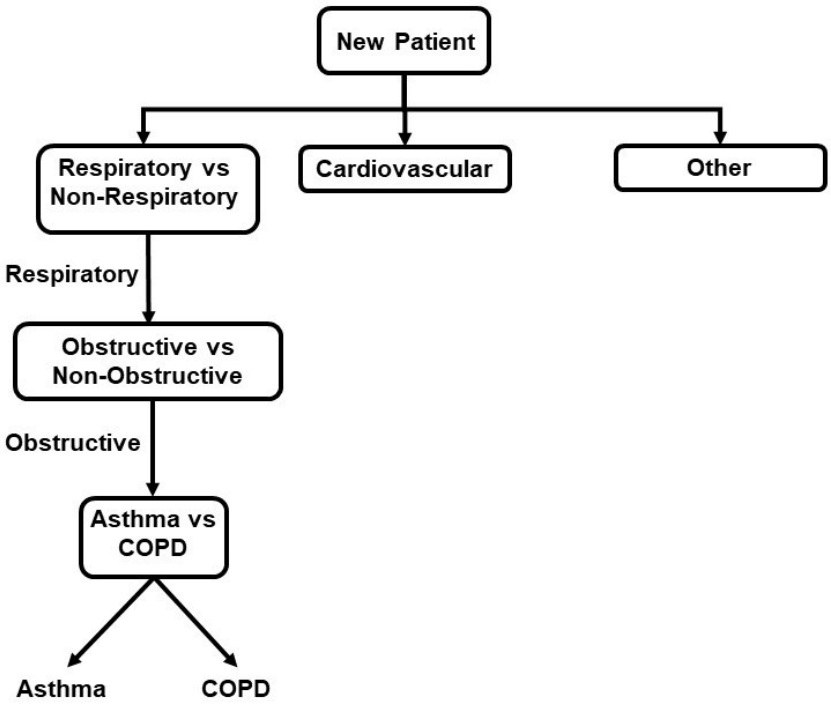

Figure 3-4: Classification scheme used in cough sound analysis

3.6.3 Discussion

While other types of clustering analysis are possible, the results from our simple k-means cluster analysis shown in Table 3.4 and Table 3.5 do not reveal any clear clusters in this set of features that map to disease diagnosis.

However, there are some results worth noting. Cluster 1 contained relatively few instances of COPD patients. This might indicate two difficulties: 1) differentiating patients with asthma from those with allergic rhinitis (since they appear in equal quantities in this cluster), and 2) differentiating asthma/allergic rhinitis patients from healthy patients. Additionally, the two features defining this cluster were Rate

Feature	Cluster 1	Cluster 2	Cluster 3
Rate of Decay	1.01	-0.47	-1.01
Zero Crossing Irregularity	-0.06	0.08	-0.12
Variance	-0.33	-0.35	2.12
Kurtosis	1.00	-0.49	-0.91
Number of Patients	29	11	39

Table 3.5: Summary of normalized feature values from k-means cluster analysis

of Decay and Kurtosis. Figure 3-5 illustrates the difficulty of determining pulmonary health when using these two features.

Cluster 2, though containing the smallest quantity of patients, is mostly composed of asthma and allergic rhinitis patients. When analyzing the average feature values for this cluster, we noted that the main difference came from the Zero Crossing Irregularity feature. Figure 3-6 shows the difficulty of detecting Obstructive vs. Non-obstructive pulmonary disease when using Zero Cross Irregularity and Rate of Decay.

Cluster 3 had many examples of all disease types, and no clear insight can be gained from it.

3.7 Supervised Analysis (Trained with Non-Comorbid and Comorbid Patients)

3.7.1 Methods

All of the cough features were treated as continuous variables and standardized to have zero-mean and unit-variance. Logistic regression models with L1-penalty were used to create the binary classifiers. 70 percent of the data were used for training,

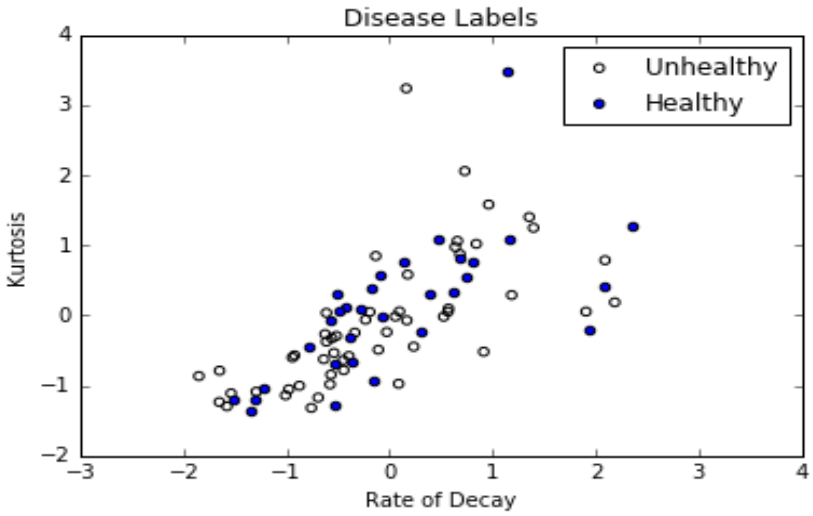

Figure 3-5: Plot of patients (Unhealthy vs. Healthy) using Kurtosis and Rate of Decay

30 percent for testing. During every trial, the data split for testing and training was done randomly. 100 training trials were run. To determine the ideal penalization parameter during each training trial, 100 trials of randomized cross validation were run over an exponential distribution; the parameter value with the lowest validation error was chosen to create the final tested classifier.

3.7.2 Results (Coughs in Isolation)

We first analyzed the utility of cough sounds in isolation.

Wet vs. Dry Cough

Figure 3-7 shows the distribution of coughs when comparing kurtosis and rate of decay, while Table 3.6 shows the results from our classifier for wet/dry coughs. It demonstrates that our classifier has high specificity (0.87) and sensitivity (1.0). The

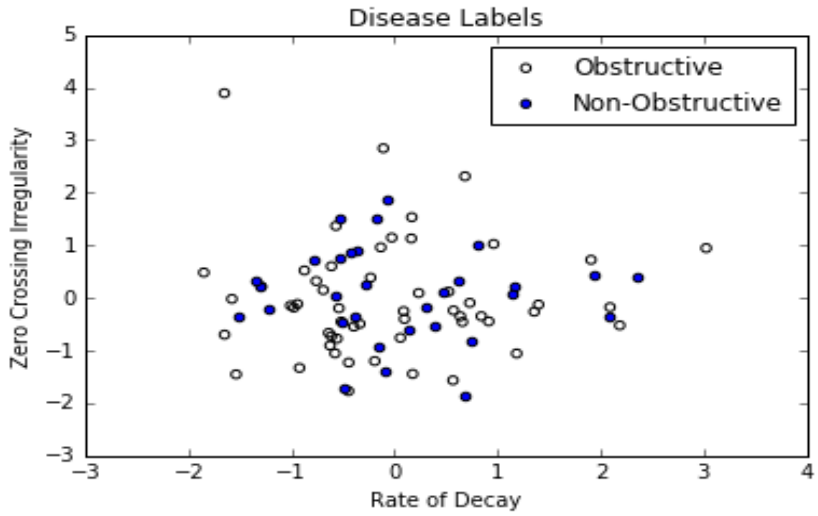

Figure 3-6: Plot of paatients (Obstructive vs. Non-obstructive) using Zero Cross Irregulairty and Rate of Decay

classifier also attained an AUC of 0.94. Figure 3-8 shows the classifier's ROC curve. These metrics indicate that our features are suitable for detecting wet/dry coughs.

Other Pulmonary Classifiers

The first classifier (Figure 3-9) determined pulmonary health, without specifying a disease (Healthy vs. Unhealthy). All of the data were used when creating this

	Percentiles (25th - 50th - 75th)
Accuracy	0.85 - **0.89** - 0.92
AUC	0.91 - **0.94** - 0.96
Sensitivity	1.00 - **1.00** - 1.00
Specificity	0.78 - **0.87** - 0.96

Table 3.6: Wet/Dry Classifier Results

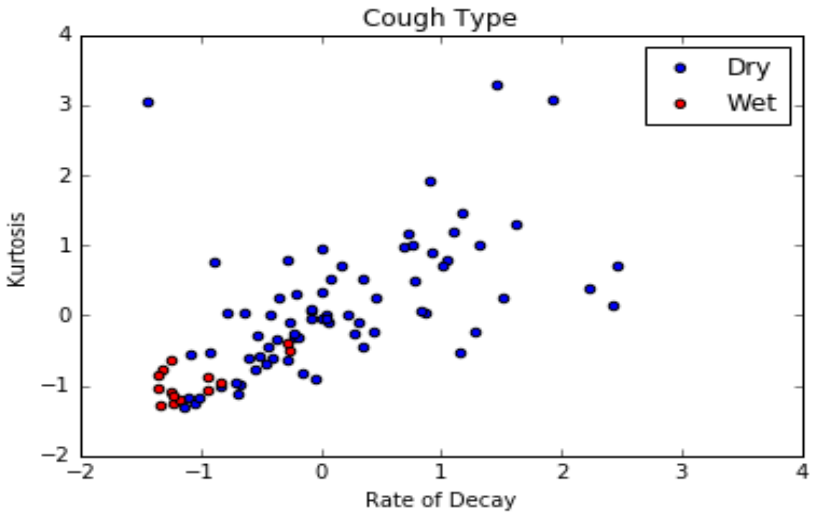

Figure 3-7: Plot of wet/dry coughs comparing the features kurtosis and rate of decay

Classifier	**AUC**
Healthy vs. Unhealthy	0.65 - **0.74** - 0.80
Obstructive vs. Non-obstructive	0.70 - **0.75** - 0.79
COPD vs. Asthma	0.75 - **0.81** - 0.83

Table 3.7: Performance of classifiers utilizing cough features in isolation

classifier.

The second classifier (Figure 3-10) determined whether patients had an obstructive pulmonary disease (COPD or asthma) or a non-obstructive pulmonary disease (allergic rhinitis). Healthy patients were omitted when creating this classifier.

The third classifier (Figure 3-11) determined whether patients had COPD or asthma. Healthy and allergic rhinitis patients were omitted when creating this classifier.

Table 3.7 summarizes the performance of the three classifiers using only the cough features.

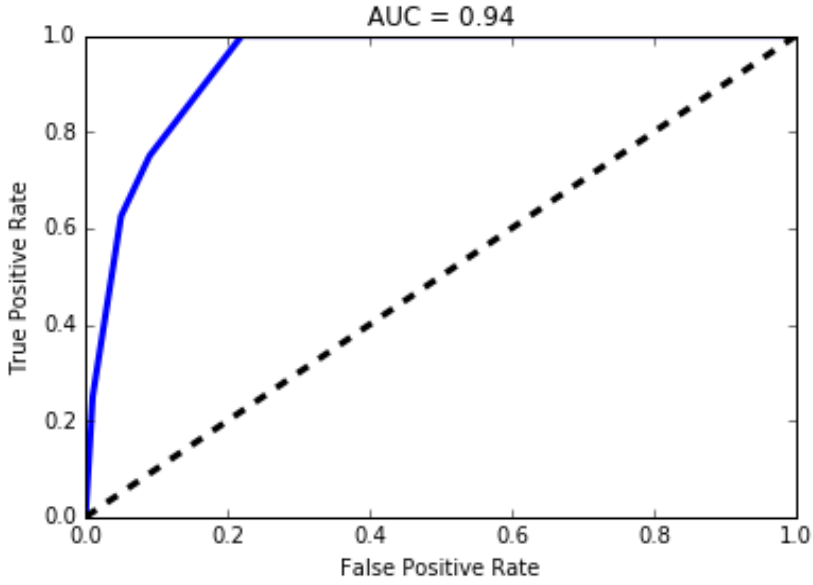

Figure 3-8: Receiver operator curve for wet-dry cough classifier

3.7.3 Results (Coughs Integrated into Mobile Kit)

The classifiers (Healthy vs. Unhealthy, Obstructive vs. Non-obstructive, COPD vs. Asthma) were retrained 14 times. The new trials used the peak flow meter, questionnaire, and lung sound features in isolation, all pairings between peak flow meter, questionnaire, lung sounds, and cough features, and all four sets combined. The same procedures described in the previous section were followed here.

Tables 3.8, 3.9, and 3.10 show the median AUC for each classification (Healthy vs. Unhealthy, Obstructive vs. Non-obstructive, and COPD vs. Asthma, respectively) and for different combination of diagnostic tools (L=lung sounds; Q=questionnaire, P=peak flow meter). Results are shown with and without the addition of cough analysis features.

Feature Set(s)	AUC without Cough Features	AUC with Cough Features
L	0.57 - **0.61** - 0.64	0.66 - **0.71** - 0.76
Q	0.94 - **0.98** - 0.99	0.95 - **0.97** - 0.99
P	0.81 - **0.89** - 0.94	0.80 - **0.86** - 0.91
Q + P	0.93 - **0.96** - 0.99	0.92 - **0.96** - 0.99
Q + L	0.95 - **0.97** - 1.00	0.95 - **0.97** - 1.00
P + L	0.86 - **0.91** - 0.94	0.80 - **0.88** - 0.95
Q + P + L	0.94 - **0.96** - 0.99	0.94 - **0.97** - 0.99

Table 3.8: Performance of Healthy vs. Unhealthy classifier utilizing different feature set combinations

Feature Set(s)	AUC without Cough Features	AUC with Cough Features
L	0.56 - **0.60** - 0.64	0.69 - **0.73** - 0.80
Q	0.76 - **0.86** - 0.91	0.73 - **0.79** - 0.88
P	0.88 - **0.92** - 0.95	0.80 - **0.87** - 0.92
Q + P	0.78 - **0.84** - 0.93	0.77 - **0.86** - 0.93
Q + L	0.75 - **0.81** - 0.88	0.75 - **0.77** - 0.88
P + L	0.84 - **0.91** - 0.96	0.79 - **0.89** - 0.95
Q + P + L	0.81 - **0.88** - 0.94	0.79 - **0.88** - 0.93

Table 3.9: Performance of Obstructive vs. Non-obstructive classifier utilizing different feature set combinations

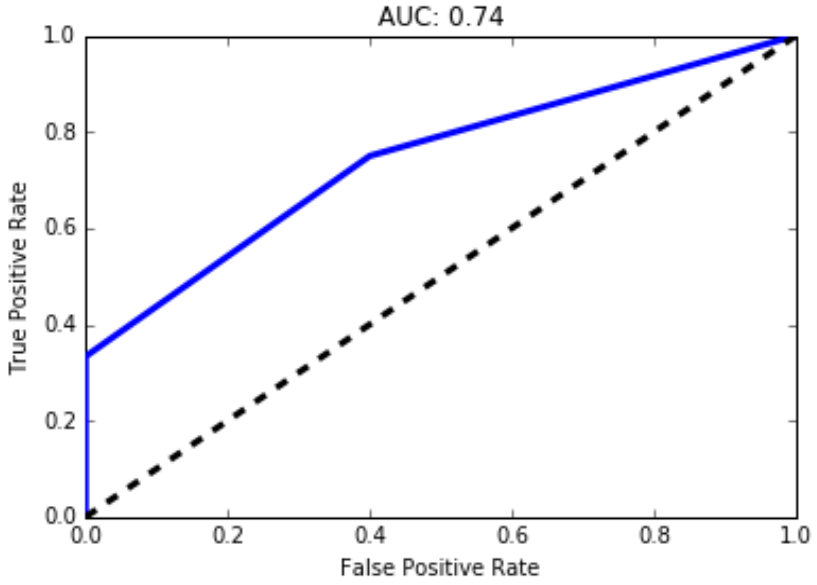

Figure 3-9: Receiver operator curve for Healthy vs. Unhealthy classifier using cough features in isolation

Coefficient Analysis

Figures 3-12, 4-6, and 3-14 note the top 10 coefficients of the pulmonary health, obstructive pulmonary disease, and COPD/asthma classifiers, respectively.

3.7.4 Discussion

Wet vs. Dry Cough

Our classification algorithm for cough type was able to accurately distinguish between wet and dry coughs using only a few time-domain features. This contrasts with recent attempts at creating a wet/dry classifier, which used many more features (both in time- and frequency-domain) and more complicated algorithms (like neural networks) [27, 29]. This is promising for two reasons. First is cost–coughs can be

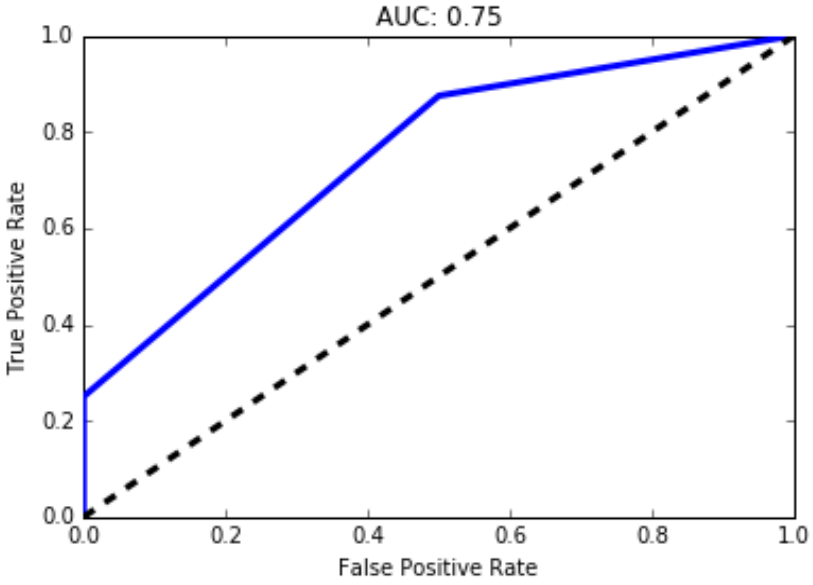

Figure 3-10: Receiver operator curve for Obstructive vs. Non-obstructive classifier using cough features in isolation

recorded directly with the phone, eliminating the need for the Mobile Kit's custom-made stethoscope. Second is simplicity–coughs alone do not require the use of extra peripherals, like the peak flow meter.

Additionally, the questionnaire does not need to be completed. This is particularly useful because there are groups which might have trouble with it–children, for example, may not be able to answer all of the questions.

Our analysis demonstrates that the detection of coughs, which we have shown are highly specific at detecting unhealthy patients, can be automated through our classifier.

A key limitation of this analysis is the small number of wet coughs. While we are optimistic about our results, we are unable to predict with a high degree of certainty that our algorithm will generalize well. Another iteration of this analysis should be

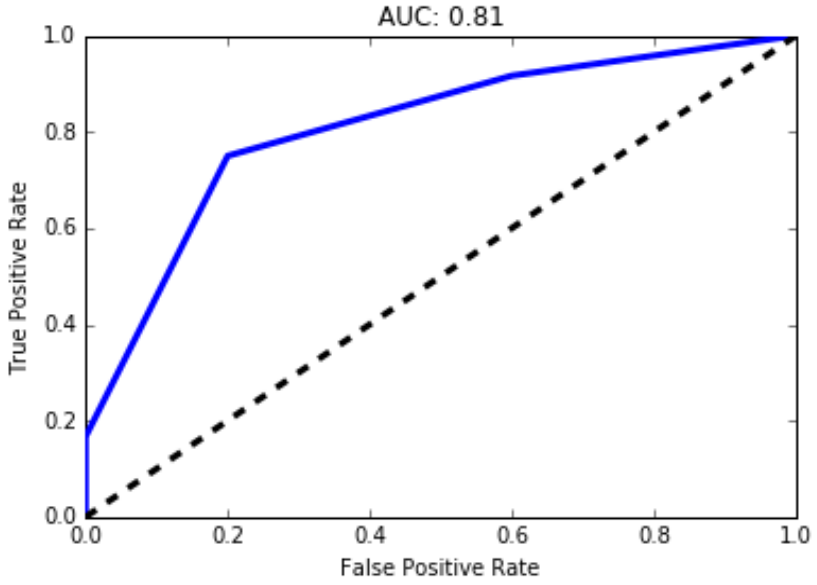

Figure 3-11: Receiver operator curve for COPD vs. Asthma classifier using cough features in isolation

conducted, with a greater sample of wet coughs.

Pulmonary Health Screening

The questionnaire by itself is able to provide near-perfect classification of pulmonary health (median AUC: 0.98). Even though cough features have been previously shown to be effective at detecting whether a cough is wet or dry, by themselves these features have moderate performance for pulmonary disease detection (median AUC: 0.74). No combination of tools (including cough features) performs better than the questionnaire in isolation.

Generally, cough features do not have an added benefit as part of a pulmonary health screener. However, they do increase performance when integrated with lung sounds (median AUC increases from 0.61 to 0.71). It should be noted, though, that

Feature Set(s)	AUC without Cough Features	AUC with Cough Features
L	0.57 - **0.64** - 0.75	0.75 - **0.83** - 0.92
Q	0.86 - **0.93** - 0.96	0.80 - **0.90** - 0.95
P	0.83 - **0.96** - 1.00	0.80 - **0.85** - 0.95
Q + P	0.79 - **0.88** - 0.92	0.80 - **0.82** - 0.95
Q + L	0.89 - **0.93** - 0.97	0.85 - **0.90** - 0.95
P + L	0.79 - **0.92** - 0.96	0.80 - **0.88** - 0.95
Q + P + L	0.82 - **0.92** - 1.00	0.75 - **0.80** - 0.90

Table 3.10: Performance of COPD vs. Asthma classifier utilizing different feature set combinations

cough features alone achieve better performance.

In the coefficient analysis of the pulmonary health classifier, cough signal variance was the fifth most important feature. However, all coefficients beyond the top three have values close to zero (when considering their medians over 100 trials). This follows the above finding that overall cough features do not improve the performance of the Mobile Kit.

Obstructive Pulmonary Health Screening

When considering feature sets in isolation, the peak flow meter performed best when screening for obstructive disease (median AUC: 0.92). Out of all possible combinations, the best performance was achieved from the peak flow meter in isolation. Even though the questionnaire decreases in performance when compared to the pulmonary health classifier, it still performs fairly well (median AUC: 0.86). Just as in the pulmonary health classifier, the classifier trained on lung sound features increased in performance when paired with cough features (median AUC increased from 0.60 to 0.73), though the cough features in isolation achieved slightly better performance

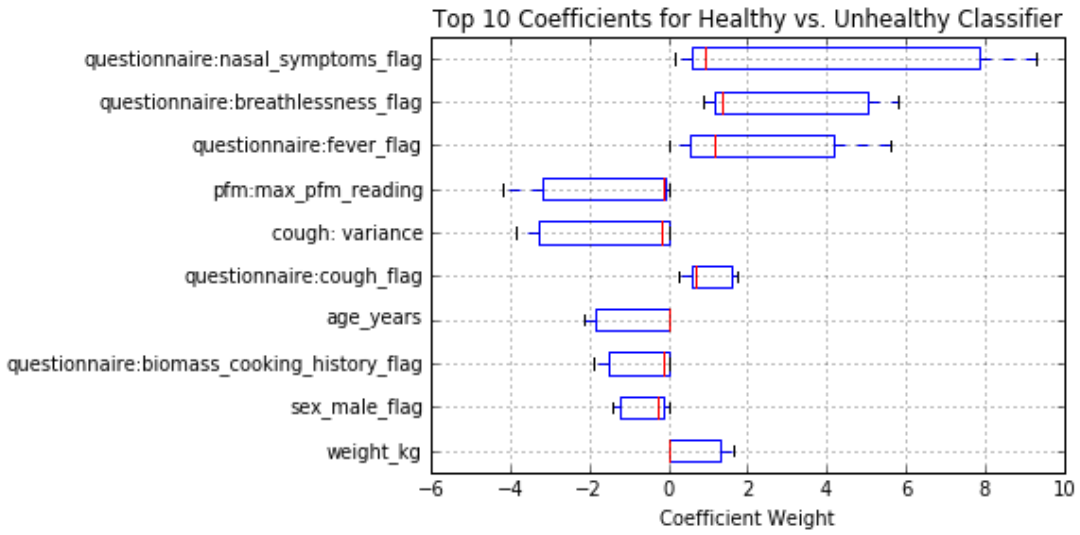

Figure 3-12: Top 10 coefficients of pulmonary health classifier when utilizing features extracted from all mobile kit tools

(median AUC: 0.75).

The coefficient analysis for the obstructive disease classifier reinforces the above findings. The questionnaire and peak flow meter make up all but one of the top 10 features. Zero cross irregularity (a cough feature) was the sixth most important feature, but again, like in the case of pulmonary health, its median value over 100 trials was near zero.

Asthma and COPD Screening

When considering feature sets in isolation, the peak flow meter again achieved the best performance (median AUC: 0.96). The peak flow meter in isolation also performed best when compared to all possible combinations, though the questionnaire had comparable results (median AUC: 0.93). Additionally, the classifier trained on lung sound features increased in performance when paired with cough features

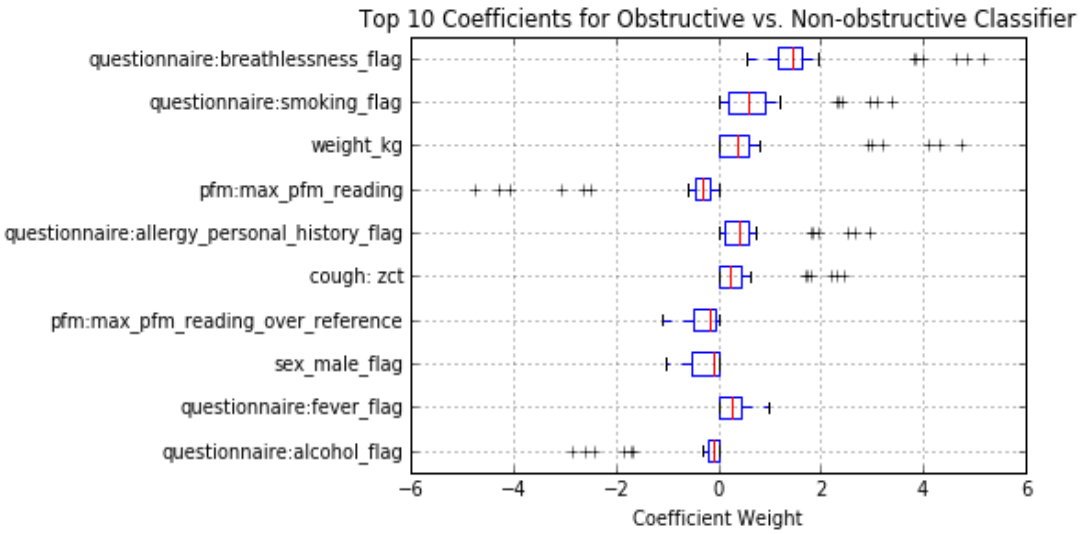

Figure 3-13: Top 10 coefficients of obstructive pulmonary disease classifier when utilizing features extracted from all mobile kit tools

(median AUC increased from 0.64 to 0.83). Unlike the other two classifiers, this combination achieved slightly better performance than cough features in isolation (median AUC: 0.81).

Unlike the pulmonary health and obstructive disease classifiers, which had median coefficient weights that did not deviate greatly from 0, the COPD/Asthma classifier had a large variance among its coefficient weights. The two strongest features, kurtosis (a cough feature) and the maximum peak flow meter reading, reinforce the finding that in combination these tools create a high performing classifier.

Utility of Cough Sound Features

Cough features in isolation did not achieve high performance for any of the classifiers, although it consistently performed better than lung sound features. The peak flow meter and the questionnaire were the top tools for all three classifiers. There was no

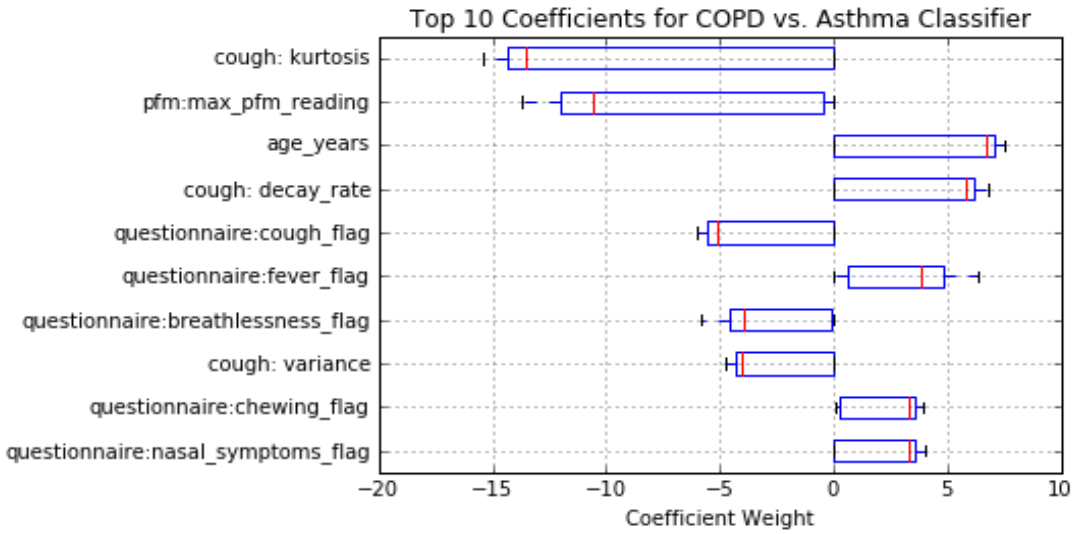

Figure 3-14: Top 10 coefficients of COPD/asthma classifier when utilizing features extracted from all mobile kit tools

clear improvement to their performance when combined with cough features. Overall, even though cough features in isolation can provide moderate performance (median AUC ranging from 0.74 to 0.81), we are not confident that they can form an integral part of the screening and diagnosis of pulmonary disease.

3.8 Supervised Analysis (Trained with Non-Comorbid Patients)

3.8.1 Methods

We repeated the supervised analysis, but trained the classifiers only with pure (non-comorbid) patients, and tested the classifiers with both pure and comorbid patients. The motivation for doing so was to see if training classifiers on pure patients achieved

better results. We hypothesized that it would, since classifiers trained on pure patients might learn more defining characteristics of each disease. Other than this distinction about what data were used for training, the same methods from the previous supervised analysis were followed here.

3.8.2 Results

Pulmonary Health

The results of training the pulmonary health classifier using only non-comorbid patients are shown in Table 3.11. The results of the same classifier, but trained using both comorbid and non-comorbid patients, have also been included for ease of comparison. On average, the classifier trained using non-comorbid patients achieved an increase of 2.13% in median AUC, and a decrease of 4.47% in the AUC's interquartile range.

Obstructive Pulmonary Disease

The results of training the obstructive pulmonary disease classifier using only non-comorbid patients are shown in Table 3.12. The results of the same classifier, but trained using both comorbid and non-comorbid patients, have also been included for ease of comparison. On average, the classifier trained using non-comorbid patients achieved an an increase of 3.4% in median AUC, and an increase of 1.93% in the AUC's interquartile range.

COPD/Asthma

The results of training the COPD/Asthma classifier using only non-comorbid patients are shown in Table 3.13. The results of the same classifier, but trained using both

Features Used	**AUC (trained on ALL patients)**	**AUC (trained on PURE patients)**	**Delta (+/- %)**
L	0.57 - **0.61** - 0.64	0.60 - **0.61** - 0.63	+ 0
Q	0.94 - **0.98** - 0.99	1.00 - **1.00** - 1.00	+ 2
P	0.81 - **0.89** - 0.94	0.90 - **0.92** - 0.94	+ 3
C	0.65 - **0.74** - 0.80	0.63 - **0.69** - 0.74	- 5
L + C	0.66 - **0.71** - 0.76	0.62 - **0.69** - 0.73	- 2
Q + C	0.95 - **0.97** - 0.99	0.99 - **1.00** - 1.00	+ 3
P + C	0.80 - **0.86** - 0.91	0.85 - **0.90** - 0.94	+ 4
Q + P	0.93 - **0.96** - 0.99	1.00 - **1.00** - 1.00	+ 4
Q + L	0.95 - **0.97** - 1.00	1.00 - **1.00** - 1.00	+ 3
P + L	0.86 - **0.91** - 0.94	0.92 - **0.94** - 0.97	+ 3
Q + P + C	0.92 - **0.96** - 0.99	0.99 - **1.00** - 1.00	+ 4
Q + L + C	0.95 - **0.97** - 1.00	1.00 - **1.00** - 1.00	+ 3
P + L + C	0.80 - **0.88** - 0.95	0.86 - **0.91** - 0.94	+ 3
Q + P + L	0.94 - **0.96** - 0.99	1.00 - **1.00** - 1.00	+ 4
Q + P + L + C	0.94 - **0.97** - 0.99	0.99 - **1.00** - 1.00	+ 3

Table 3.11: Performance of Unhealthy classifier trained on all patients (comorbid and non-comorbid) vs. trained on pure patients (non-comorbid)

Features Used	**AUC (trained on ALL patients)**	**AUC (trained on PURE patients)**	**Delta (+/- %)**
L	0.56 - **0.60** - 0.64	0.50 - **0.57** - 0.70	- 3
Q	0.76 - **0.86** - 0.91	0.78 - **0.86** - 0.99	+ 0
P	0.88 - **0.92** - 0.95	0.86 - **0.94** - 0.96	+ 2
C	0.70 - **0.75** - 0.79	0.75 - **0.80** - 0.88	+ 5
L + C	0.69 - **0.73** - 0.80	0.76 - **0.81** - 0.86	+ 8
Q + C	0.73 - **0.79** - 0.88	0.78 - **0.86** - 0.97	+ 7
P + C	0.80 - **0.87** - 0.92	0.82 - **0.89** - 0.93	+ 2
Q + P	0.78 - **0.84** - 0.93	0.85 - **0.94** - 0.99	+ 10
Q + L	0.75 - **0.81** - 0.88	0.80 - **0.85** - 0.98	+ 4
P + L	0.84 - **0.91** - 0.96	0.84 - **0.90** - 0.95	- 1
Q + P + C	0.77 - **0.86** - 0.93	0.84 - **0.91** - 0.98	+ 5
Q + L + C	0.75 - **0.77** - 0.88	0.80 - **0.86** - 0.93	+ 9
P + L + C	0.79 - **0.89** - 0.95	0.84 - **0.89** - 0.93	+ 0
Q + P + L	0.81 - **0.88** - 0.94	0.80 - **0.90** - 0.99	+ 2
Q + P + L + C	0.79 - **0.88** - 0.93	0.83 - **0.89** - 0.99	+ 1

Table 3.12: Performance of Obstructive classifier trained on all patients (comorbid and non-comorbid) vs. trained on pure patients (non-comorbid)

comorbid and non-comorbid patients, have also been included for ease of comparison. On average, the classifier trained using non-comorbid patients achieved an an increase of 2.73% in median AUC, and a decrease of 8.67% in the AUC's interquartile range.

Features Used	**AUC (trained on ALL patients)**	**AUC (trained on PURE patients)**	**Delta (+/- %)**
L	0.57- **0.64** - 0.75	0.53 - **0.57** - 0.60	- 7
Q	0.86 - **0.93** - 0.96	0.93 - **0.95** - 0.97	+ 2
P	0.83 - **0.96** - 1.00	0.92 - **0.94** - 0.96	- 2
C	0.75 - **0.81** - 0.83	0.73 - **0.76** - 0.80	- 5
L + C	0.75 - **0.83** - 0.92	0.70 - **0.74** - 0.79	- 9
Q + C	0.80 - **0.90** - 0.95	0.93 - **0.96** - 0.97	+ 6
P + C	0.80 - **0.85** - 0.95	0.87 - **0.90** - 0.93	+ 5
Q + P	0.79 - **0.88** - 0.92	0.91 - **0.96** - 0.98	+ 8
Q + L	0.89 - **0.93** - 0.97	0.94 - **0.96** - 0.98	+ 3
P + L	0.79 - **0.92** - 0.96	0.84 - **0.92** - 0.95	+ 0
Q + P + C	0.80 - **0.82** - 0.95	0.88 - **0.95** - 0.97	+ 13
Q + L + C	0.85 - **0.90** - 0.95	0.94 - **0.96** - 0.98	+ 6
P + L + C	0.80 - **0.88** - 0.95	0.87 - **0.89** - 0.92	+ 1
Q + P + L	0.82 - **0.92** - 1.00	0.93 - **0.97** - 0.98	+ 5
Q + P + L + C	0.75 - **0.80** - 0.90	0.91 - **0.95** - 0.96	+ 15

Table 3.13: Performance of COPD/Asthma classifier trained on all patients (comorbid and non-comorbid) vs. trained on pure patients (non-comorbid)

3.8.3 Discussion

All classifier results improved when trained with only non-comorbid patients. However, it should be noted that when the classifiers used features from lung sounds and/or cough sounds, and were trained with non-comorbid paitents, they tended to perform worse. Additionally, the pulmonary health and COPD/asthma classifiers tended to achieve a smaller interquartile range for AUC.

Physiologically, the better performance from classifiers trained on non-comorbid patients makes sense – these are "pure" patients, with respect to disease. It is much more likely that the features extracted from these patients' data are more indicative of their respective disease. Training classifiers on both comorbid and non-comorbid patients introduces a degree of uncertainty into the classifier. For example, a patient with both COPD and allergic rhinitis might have symptoms that are more indicative of COPD than allergic rhinitis (or vice versa). However, a classifier trained on such a patient would not know this, and would attribute the distinctive features to both diseases. Based on our results, we believe that training classifiers using only non-comorbid patients is preferable.

It should be noted that this finding goes against the machine learning assumption that a classifier be trained from independent samples extracted from an i.i.d. space. A consequence of this is that while we have demonstrated stronger predictive power, we are not able to guarantee that our classifiers generalize well.

3.9 Future Improvements

Inclusion of Infectious Diseases

Our analyses have focused on COPD, asthma, and allergic rhinitis. While these are prevalent and burdensome diseases, we have not addressed pulmonary infections. Tuberculosis and pneumonia (especially in children) create a huge burden on the developing world. Including these diseases into our pulmonary health kit could help alleviate this healthcare concern. We have begun to address this by creating a new classification scheme, summarized in Figure 4-4.

New Cough Features

For this analysis, we chose computationally inexpensive time-domain features to extract from the cough signal. However, there are many more features we can add which can potentially improve the contribution of cough sounds to the classifiers. For example, there are many frequency-domain features which are used for speech recognition, like the Mel-Frequency Cepstral Coefficient (MFCC), entropy, and formants.

Improved Interpretability

Interpretability is a main concern in all machine learning problems involving healthcare. We have used logistic regression classifiers in order to address this, given the insight the coefficients give into what information helped the system make its decision.

In future work, we wish to include other, potentially more interpretable, models. For example, a Bayesian model (such as Bayesian logistic regression) allows for the

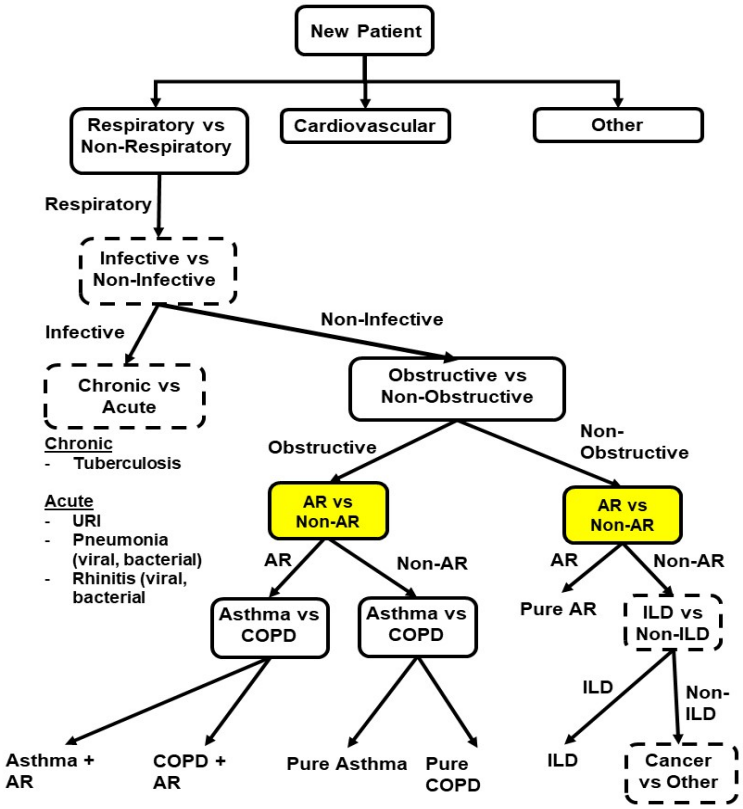

Figure 3-15: New classification scheme. Dotted lines denote classifiers that will be created in the future. Yellow boxes denote additions to the old scheme.

use of known prior distributions of unknown parameters, and provides posterior predictive distributions for all learned paramters.

Chapter 4

Analysis of Pulmonary Function Testing (PFT) Data

Pulmonary Function Testing (PFT) is the current gold-standard for pulmonary disease diagnosis. However, the accurate interpretation of its results requires training and experience. The automatic analysis of these data would allow technicians with little interpretive experience to accurately diagnose pulmonary disease. In this chapter, we search for any hidden clusters with the data extracted from PFT machines, and see how they correlate with disease and any patient subpopulations. We then conduct a variety of supervised learning analyses: first, we use various combinations of PFT machines to train the classifiers within the new diagnostic protocol; second, we analyze the use of adding clinical features (lung sounds and questionnaire) to the PFT data; third, we explore a new classification scheme composed of various independent disease classifiers. We conclude by comparing the performance of the classifiers created from the PFT data to those results from the Mobile Kit data.

4.1 Motivation

Pulmonary Function Testing (PFT) requires interpretation of results by an expert, using prior knowledge and guidelines established by the European Respiratory Society (ERS) and the American Thoracic Society (ATS), usually in the form of hard cut-offs of measured physiological parameters. Recently, there has been a growing interest in the role artificial intelligence can play in improving the accuracy of pulmonary disease diagnosis from PFT data.

4.1.1 Past Work

A 2012 study analyzed the data acquired from spirometry and forced oscillation machines on 50 subjects (25 healthy, 25 with COPD). They concluded that k-nearest neighbors (kNN), support vector machines (SVM), and artificial neural networks (ANN) provided good results (AUC $>$ 0.95) [60].

A 2017 study analyzed PFT data from 968 subjects. They developed an automated algorithm using the ATS and ERS guidelines, and compared it to a classifier developed using machine learning. While they were able to achieve moderate results for the detection of COPD (74% accuracy), overall they achieved poor results for other diseases (asthma, interstitial lung disease, and neuromuscular disorder) with an overall accuracy of 38%. Nonetheless, the researchers from this study conclude that their classifier improves the interpretability of current diagnosis methods [59].

4.2 Machine Overview

For this analysis, we collected data from four machines, all of the MasterScreen model, manufactured by Jaegar. LabManager v5.32.0 was used to extract and com-

pile the machines' readings. An explanation of each machine, its testing procedure, and the fields extracted are summarized below.

- **Spirometry**. For this test, the patient must inhale to maximum capacity, and exhale as quickly as possible into a tube. The test measures lung capacity, as well as the rate of exhalation. For new patients, this test can take up to 30 minutes (due to the need to complete multiple training trials in order to perform the test correctly); more experienced patients generally complete the test in 15 minutes. Only one trial is done, consisting of a pre-test (before a bronchodilator is administered) and a post-test (after a bronchodilator is administered). A summary of the spirometry fields used in this analysis can be found in Table 4.1.

- **Diffusing capacity of the lungs for carbon monoxide (DLCO)**. For this test, the patient breathes in, holds his or her breath for 10 seconds, and then rapidly exhales into a gas analyzer machine. This test measures the diffusion of gas between the patient's air sacs (alveoli) and the patient's blood. The total time for this test (two trials) is eight to nine minutes. Each trial consists of a pre-test (before a bronchodilator is administered) and a post-test (after a bronchodilator is administered). A summary of the DLCO fields used in this analysis can be found in Table 4.1.

- **Body plethysmography, or Body box (BB)**. For this test, the patient sits inside an airtight box which measures tiny displacements of air in order to measure the different lung volumes inside the patient's body. The patient breathes through a mouthpiece while wearing a nose clip. The patient is asked to breathe normally; then, a shutter inside the mouthpiece closes. Measurements are taken while the patient attempts to inhale. Throughout the test, a shutter on the air tube opens and

closes five times. A summary of the BB fields used in this analysis can be found in Table 4.2.

- **Impulse oscillometry (IOS)**. This machine measures the frequency response of a patient's lungs, measuring both the real and imaginary components of the impedence. For this test, the patient places his/her mouth on a mouthpiece and air tube and breathes normally. A machine produces pulses of air which are superimposed on the patient's normal breathing. Pulmonary resistance, reactance, and impedance are recorded when waves at various frequencies (5, 10, and 20 Hz) are superimposed on normal breathing. The total time for this test (three trials) is five minutes. Unlike the other machines, IOS does not involve pre- and post-tests. A summary of the IOS fields used in this analysis can be found in Table 4.2.

4.3 Data Collection

As per the study protocol, all patients (healthy and unhealthy) were requested to undergo PFT. However, many opted out, usually out of discomfort (many found the procedure physically demanding). Table 4.3 summarizes how many patients completed the tests for each machine, in isolation and combination, as well as some basic demographic information about each group.

Field	Unit	Machine	Explanation
FVC	L	Spirometry	Forced vital capacity. The volume of air that the patient exhaled.
FEV 1	L	Spirometry	Forced expiratory volume in 1 second. The volume of air that the patient exhaled in 1 second.
FEV 1 / max VC	%	Spirometry	The ratio of FEV 1 to FVC.
PEF	L/s	Spirometry	Peak expiratory flow. The maximum flow achieved during exhalation.
MMEF 75 % 25	L/s	Spirometry	Maximal mid-expiratory flow. The maximum of expiratory flow between 25% and 75% of FVC.
TLC-SB	L	DLCO	Total lung capacity in a single breath.
RV % TLC	%	DLCO	The ratio of residual volume to total lung capacity.
DLCO-SB	mmol/(s*kPa)	DLCO	DLCO for a single breath.
DLCOc	mmol/(s*kPa)	DLCO	DLCO adjusted to the patient's hemoglobin level. Equation found [REF].
KCO	mmol/(s*kPa*L)	DLCO	The ratio of DLCO to alveolar volume.

Table 4.1: Summary of PFT fields which were included in the analysis

Field	Unit	Machine	Explanation
R5Hz	kPa/(L/s)	IOS	Resistance at 5 Hz.
R10Hz	kPa/(L/s)	IOS	Resistance at 10 Hz.
R20Hz	kPa/(L/s)	IOS	Resistance at 20 Hz.
X5Hz	kPa/(L/s)	IOS	Reactance at 5 Hz.
SG total	1/(kPa*s)	BB	Specific airway resistance over total flow range.
SG 0.5	1/(kPa*s)	BB	Specific airway resistance at flow rate equal to +/- 0.5 L/s.
TLC	L	BB	Total lung capacity.
WoB	kPa*L	BB	Work of breathing.
RV % TLC	%	BB	The raio of residual volume to total lung capacity.

Table 4.2: Summary of PFT fields which were included in the analysis

Machines	#	Age (yrs.)	Male	Weight (kg)	COPD	Asthma	AR	Healthy
S	253	47.59	140	61.00	39	96	82	73
BB	260	47.58	148	61.12	41	97	85	79
IOS	269	47.97	151	61.33	42	98	86	79
DLCO	227	46.18	129	61.54	28	85	72	73
S + BB	247	47.30	138	61.04	37	96	81	73
S + IOS	249	47.40	139	61.10	37	95	82	73
S + DLCO	216	45.98	121	61.25	26	83	69	68
BB + IOS	257	47.47	147	61.18	40	96	85	79
BB + DLCO	220	45.84	125	61.60	26	84	71	73
IOS + DLCO	223	45.95	128	61.74	26	84	72	73
S + IOS + DLCO	214	45.92	121	61.37	25	83	69	68
S + BB + IOS	244	47.18	137	61.10	36	95	81	73
S + BB + DLCO	212	45.84	119	61.21	25	83	68	68
BB + IOS + DLCO	218	45.78	125	61.71	25	84	71	73
S + BB + IOS + DLCO	210	45.78	119	61.33	24	83	68	68

Table 4.3: Summary statistics of patients who completed various combinations of the PFT machines

4.4 Data Pre-processing and Feature Selection

Each subsection below describes how the final classification features were calculated from each machine's extracted fields. All of the final features were standardized to have zero-mean and unit-variance.

4.4.1 Spirometry and DLCO

All of the utilized fields for spirometry and DLCO have positive predicted values. These predicted values were used to create the final classification features for these machines. For each field, three features were calculated, as follows:

- Pre-test (median) / predicted value
- Post-test (median) / predicted value
- [Post-test (median) - Pre-test (median)] / predicted value

4.4.2 Impulse Oscillometry

Pulmonary resistance at 5 Hz, 10 Hz, and 20 Hz (labeled as R5Hz, R10Hz, and R20Hz, respectively) have positive predicted values. For each of these three fields, three features were calculated in the same manner as for spirometry and DLCO, repeated here for convenience:

- Pre-test (median) / predicted value
- Post-test (median) / predicted value
- [Post-test (median) - Pre-test (median)] / predicted value

Pulmonary reactance at 5 Hz (X5Hz) can have negative, positive, or zero predicted value. Unlike the three IOS fields above, we could not divide by the predicted value to determine X5Hz's final three classification features. Therefore, subtraction was used, as follows:

- Pre-test (median) - predicted value
- Post-test (median) - predicted value
- [Post-test (median) - Pre-test (median)] - predicted value

4.4.3 Body Plethysmography

Body plethysmography requires five trials. The first three trials determine pulmonary conductance (or resistance). The final three trials determine lung volume. The third trial is common between the two measurements.

This exam is physically demanding, and some patients were not able to complete all trials. The data for any patient who did not complete at least one of the first three trials, and at least one of the final three trials, were excluded from the analysis.

All of the utilized fields for body plethysmography have positive predicted values, determined by [REF], except for Work of Breathing (which does not have any predicted value). When determining the median values below, only the first three trials were considered for pulmonary resistance fields (SG total and SG 0.5), while only the final three trials were considered for pulmonary volume fields (TLC and RV % TLC). When determining the median, missing trials were ignored. For each field except Work of Breathing, three features were calculated, as follows:

- Pre-test (median) / predicted value
- Post-test (median) / predicted value

- [Post-test (median) - Pre-test (median)] / predicted value

A third feature, labeled Work of Breathing, was defined as the median value of the first three trials (excluding any missing trials) was directly used as a classification feature.

4.5 Unsupervised Learning Analysis

4.5.1 Choice of Methods

We performed an unsupervised analysis to search for any disease clusters within the dataset. Additionally, we searched for any patient subpopulations.

Given the unbalanced class sizes of our data, we chose the DBSCAN algorithm which is optimized to analyze changes in the data density in multi-dimensional space.

As an alternative, the standard k-means algorithm was also used, which attempts to identify clusters based on Euclidean distance while assuming even class sizes.

4.5.2 Cluster Analysis Results

Results from DBSCAN Algorithm

The DBSCAN algorithm was unable to converge, labeling all data points as belonging to the same cluster. This algorithm assumes clusters of high-density separated by regions of low-density. Given the failure of the DBSCAN algorithm to converge, we suspect that our data are fairly evenly and sparsely distributed with no regions of high density. In order to visualize the high-dimensional data, we plotted several two-dimensional slices (plotting patients against two features). These are shown in

Figures 4-1, 4-2, and 4-3.

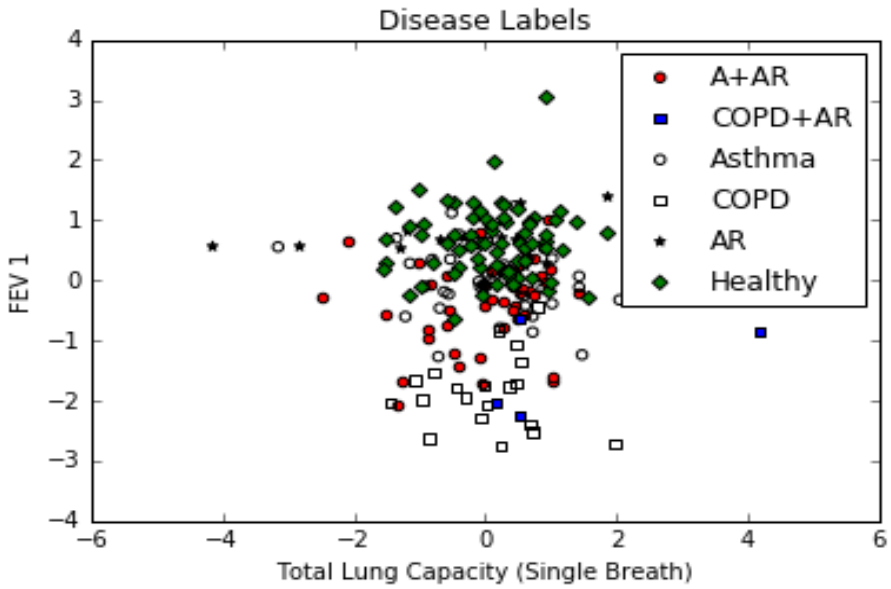

Figure 4-1: Top 10 coefficients of pulmonary health classifier when utilizing features extracted from all PFT machines

Results from K-means Algorithm

After poor results from the DBSCAN algorithm, we utilized the common k-means algorithm. For the analysis of the full PFT dataset, the maximum silhouette score suggested the use of five clusters. For the analysis of the COPD patients within the PFT dataset, the maximum silhouette score suggested the use of four clusters; however, we used two in order to combine clusters with few patients. For the analysis of the asthma patients within the PFT dataset, the maximum silhouette score suggested the use of four clusters.

All Disease Types

We began by analyzing the entire data set for any disease clusters. Table 4.4 sum-

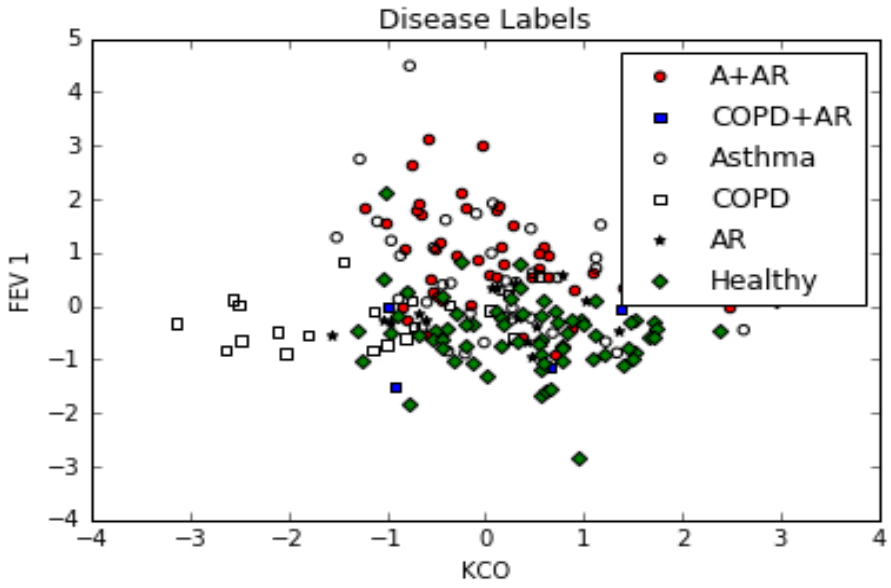

Figure 4-2: Top 10 coefficients of pulmonary health classifier when utilizing features extracted from all PFT machines

marizes the disease breakdown of each cluster.

	Healthy	**Asthma**	**AR**	**COPD**	**Total**
Cluster 1	39	11	14	0	67
Cluster 2	0	18	15	1	19
Cluster 3	3	39	22	2	49
Cluster 4	26	10	13	2	49
Cluster 5	0	5	4	19	26

Table 4.4: Intra-cluster disease breakdown within PFT dataset

COPD Patients

Table 4.5 summarizes the risk factors and symptoms of the patients belonging to the two clusters.

Asthma Patients

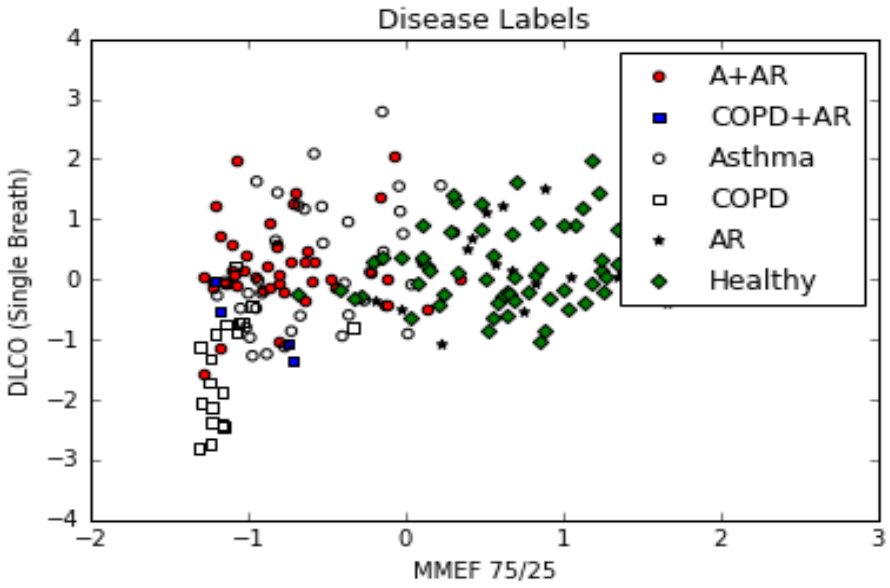

Figure 4-3: Top 10 coefficients of pulmonary health classifier when utilizing features extracted from all PFT machines

Table 4.6 summarizes the risk factors and symptoms of the patients belonging to the four clusters.

	Cluster 1 (n=12)	**Cluster 2 (n=12)**
Male (%)	75.00	83.33
Age (years)	55.33	50.94
Weight (kg)	58.61	57.57
Breathless (%)	100.00	100.00
Cough (%)	75.00	100.00
Chest Pain (%)	16.67	0.00
Fever (%)	8.33	0.00
Nasal (%)	75.00	66.67
COPD (Family History) (%)	0.00	0.00
Allergy (Family History) (%)	25.00	8.33
Allergy (Personal History) (%)	16.67	0.00
Biomass Cooking (%)	8.33	8.33
Smoking (%)	66.67	50.00
Tobacco Chewing (%)	75.00	33.33
Alcohol (%)	41.67	0.00

Table 4.5: Results of unsupervised analysis of COPD patients from PFT dataset

	Cluster 1 (n=30)	Cluster 2 (n=11)	Cluster 3 (n=11)	Cluster 4 (n=31)
Male (%)	63.33	72.72	54.55	54.84
Age (years)	50.97	59.09	52.82	52.19
Weight (kg)	59.28	57.72	53.77	62.13
Breathless (%)	80.00	100.00	100.00	77.42
Cough (%)	73.33	72.72	81.82	70.97
Chest Pain (%)	6.67	9.01	0.00	6.45
Fever (%)	3.33	0.00	0.00	0.00
Nasal Symptoms (%)	63.33	54.55	54.55	54.84
COPD (Family History) (%)	0.00	0.00	0.00	0.00
Allergy (Family History) (%)	36.67	27.27	18.18	38.71
Allergy (Personal History) (%)	20.00	18.18	27.27	32.26
Biomass Cooking (%)	6.67	0.00	9.09	16.13
Smoking (%)	30.00	54.55	36.26	25.81
Tobacco Chewing (%)	30.00	54.55	54.55	25.81
Alcohol Intake (%)	10.00	18.18	18.18	12.9

Table 4.6: Results of unsupervised analysis of asthma paptients from PFT datset

4.5.3 Discussion of Cluster Analysis Results

The inability of the DBSCAN algorithm to find any clusters suggests that there are no clear clusters within the dataset. This is reinforced by the lack of clusters in Figures 4-1, 4-2, and 4-3. Even though a supervised analysis might provide highly accurate classifiers, the data is inherently of equal density across the N-dimensional space (N = number of features). There is no indication from the analysis of these data that there are clear clusters corresponding to disease.

The results of the k-means analysis on the entire dataset (Table 4.4) suggests that there are two types of asthma and AR patients: one which is similar to healthy subjects (Clusters 1 and 4), and another which is not (Clusters 2 and 3). This can either indicate symptomatic and asymptomatic versions of the diseases, or that the asymptomatic patients had to have medication administered to allow the patients to complete the examination. These results also indicate that asthma and allergic rhinitis are very similar, and the presence of one might indicate the presence of the other. Finally, COPD does appear as a mostly separate cluster (Cluster 5). This indicates that while the PFT data are useful for detecting COPD, they may falter at distinguishing between healthy, asthma, and AR subjects.

While the above reuslts suggest that PFT data are useful for the detection of COPD, the cluster analysis of the COPD patients (Table 4.5) indicate that there are no clear COPD subpopulations. The only apparent difference between the two groups is that Cluster 1 contains more tobacco chewers and alcohol drinkers. It is possible that this indicates a type of patient who abuses tobacco and alcohol. However, further analysis is necessary given the small sample of COPD patients in the datset.

Among the asthma patents (Table 4.6), the risk factors and symptoms are fairly

uniform. Nonetheless, Cluster 2 provides an interesting result. This cluster is more heavily male (73%), older (59 years), does not have exposure to biomass cooking, and tends to smoke more (55%). This might indicate a common type of asthma patient within the Indian population.

Overall, while there are some insights gained from this analysis, the exploration of a larger dataset using alternate algorithms might provide richer results.

4.6 Supervised Learning Analysis of PFT Data in Isolation

4.6.1 Motivation

While PFT results are usually analyzed by physicians in combination with clinical features (questionnaire, lung sounds), in this section we are interested in comparing the utility of each PFT machine. Therefore, we trained and tested the various classifiers using only various combinations of the PFT machines.

4.6.2 Classification Design

In the following analysis, we follow the new multi-layer classifier structure summarized in Figure 4-4, which has been described previously.

4.6.3 Methods

All of the features were standardized to have zero-mean and unit-variance. Logistic regression models with L1-penalty were used to create the binary classifiers. 70 percent of the data were used for training, 30 percent for testing. During every trial,

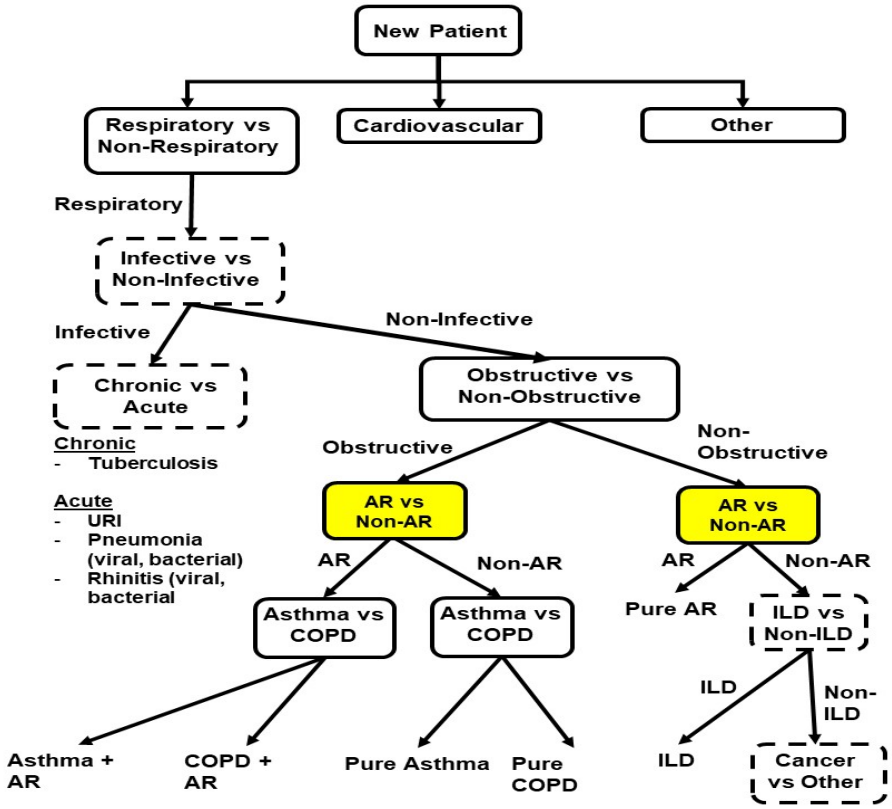

Figure 4-4: New classification scheme. Dotted lines denote classifiers that will be created in the future. Yellow boxes denote additions to the old scheme.

the data split for testing and training was done randomly. 100 training trials were run. To determine the ideal penalization parameter during each training trial, 100 trials of randomized cross validation were run over an exponential distribution; the parameter value with the lowest validation error was chosen to create the final tested classifier.

4.6.4 Results of Supervised Learning Analysis

Table 4.7 summarizes the performance of the pulmonary health classifier when trained on various combinations of the features extracted from the PFT machines. With the same training protocol, Table 4.8 summarizes the performance of the obstructive pulmonary disease classifier, Table 4.9 of the AR (obstructive) classifier, Table 4.10 of the COPD (comorbid) classifier, Table 4.11 of the COPD (non-comorbid) classifier, and Table 4.12 of the AR (non-obstructive) classifier.

PFT Machine(s) Used	**AUC**	**Sensitivity**	**Specificity**
S	0.88 - **0.90** - 0.92	0.71 - **0.76** - 0.82	0.93 - **0.93** - 1.00
BB	0.87 - **0.89** - 0.91	0.72 - **0.74** - 0.79	0.88 - **0.94** - 0.94
IOS	0.85 - **0.87** - 0.90	0.61 - **0.68** - 0.76	0.88 - **0.94** - 1.00
DLCO	0.71 - **0.77** - 0.81	0.57 - **0.68** - 0.79	0.67 - **0.73** - 0.87
S + BB	0.88 - **0.90** - 0.91	0.73 - **0.76** - 0.81	0.87 - **0.93** - 1.00
S + IOS	0.88 - **0.90** - 0.92	0.70 - **0.76** - 0.81	0.93 - **0.93** - 1.00
S + DLCO	0.88 - **0.90** - 0.93	0.69 - **0.78** - 0.84	0.86 - **0.93** - 0.93
BB + IOS	0.87 - **0.89** - 0.91	0.71 - **0.76** - 0.82	0.88 - **0.94** - 1.00
BB + DLCO	0.85 - **0.88** - 0.92	0.70 - **0.74** - 0.81	0.87 - **0.87** - 0.93
IOS + DLCO	0.87 - **0.89** - 0.91	0.66 - **0.81** - 0.88	0.73 - **0.87** - 0.93
S + BB + IOS	0.88 - **0.90** - 0.92	0.69 - **0.75** - 0.81	0.87 - **0.93** - 1.00
S + BB + DLCO	0.87 - **0.90** - 0.92	0.70 - **0.74** - 0.84	0.86 - **0.93** - 1.00
S + IOS + DLCO	0.89 - **0.91** - 0.93	0.74 - **0.77** - 0.84	0.86 - **0.93** - 1.00
BB + IOS + DLCO	0.87 - **0.89** - 0.91	0.71 - **0.77** - 0.84	0.80 - **0.87** - 0.93
S + BB + IOS + DLCO	0.88 - **0.90** - 0.92	0.70 - **0.77** - 0.83	0.86 - **0.93** - 1.00

Table 4.7: Performance of Healthy vs. Unhealthy classifier different PFT feature set combinations

PFT Machine(s) Used	AUC	Sensitivity	Specificity
S	0.91 - **0.94** - 0.96	0.78 - **0.89** - 0.96	0.82 - **0.91** - 0.91
BB	0.88 - **0.91** - 0.95	0.79 - **0.89** - 0.93	0.73 - **0.82** - 0.91
IOS	0.85 - **0.89** - 0.91	0.71 - **0.79** - 0.86	0.77 - **0.85** - 0.92
DLCO	0.64 - **0.66** - 0.71	0.43 - **0.59** - 0.78	0.50 - **0.70** - 0.80
S + BB	0.90 - **0.92** - 0.95	0.81 - **0.88** - 0.96	0.73 - **0.82** - 0.91
S + IOS	0.92 - **0.95** - 0.97	0.85 - **0.88** - 0.96	0.82 - **0.91** - 0.87
S + DLCO	0.91 - **0.94** - 0.96	0.77 - **0.86** - 0.95	0.80 - **0.90** - 1.00
BB + IOS	0.90 - **0.93** - 0.96	0.81 - **0.87** - 0.96	0.79 - **0.87** - 0.95
BB + DLCO	0.88 - **0.91** - 0.94	0.77 - **0.86** - 0.91	0.78 - **0.89** - 1.00
IOS + DLCO	0.85 - **0.88** - 0.92	0.68 - **0.77** - 0.86	0.80 - **0.90** - 1.00
S + IOS + DLCO	0.91 - **0.94** - 0.97	0.82 - **0.91** - 0.95	0.78 - **0.89** - 0.89
S + BB + IOS	0.90 - **0.93** - 0.96	0.81 - **0.92** - 0.96	0.80 - **0.80** - 0.90
S + BB + DLCO	0.90 - **0.93** - 0.96	0.77 - **0.86** - 0.91	0.78 - **0.89** - 1.00
BB + IOS + DLCO	0.88 - **0.92** - 0.95	0.77 - **0.86** - 0.91	0.78 - **0.89** - 1.00
S + BB + IOS + DLCO	0.91 - **0.93** - 0.96	0.82 - **0.91** - 0.95	0.75 - **0.88** - 1.00

Table 4.8: Performance of Obstructive vs. Non-obstructive classifier different PFT feature set combinations

Coefficient Analysis

When each of the classifiers was trained using the data from all of the PFT machines, we avareaged the top 10 coefficient weight magnitudes. The top 10 coefficients for the pulmonary health classifier are summarized in Figure 4-5, for the obstructive pulmonary disease classifier in Figure 4-6, for the AR (obstructive) classifier in Figure 4-7, for the COPD (comorbid) classifier in Figure 4-8, for the COPD (non-comorbid) classifier in Figure 4-9, and for the AR (non-obstructive) classifier in Figure 4-10

PFT Machines(s) Used	AUC	Sensitivity	Specificity
S	0.69 - **0.72** - 0.76	0.69 - 0.85 - 0.92	0.43 - 0.57 - 0.71
BB	0.63 - **0.66** - 0.68	0.54 - **0.69** - 0.85	0.45 - **0.60** - 0.73
IOS	0.65 - **0.68** - 0.72	0.46 - **0.62** - 0.77	0.53 - **0.73** - 0.80
DLCO	0.66 - **0.71** - 0.78	0.60 - **0.80** - 1.00	0.46 - **0.62** - 0.77
S + BB	0.65 - **0.67** - 0.71	0.54 - **0.77** - 0.92	0.40 - **0.57** - 0.73
S + IOS	0.66 - **0.69** - 0.84	0.46 - **0.62** - 0.85	0.46 - **0.69** - 0.85
S + DLCO	0.67 - **0.72** - 0.77	0.58 - **0.70** - 0.90	0.50 - **0.67** - 0.83
BB + IOS	0.64 - **0.67** - 0.71	0.44 - **0.62** - 0.77	0.53 - **0.73** - 0.87
BB + DLCO	0.64 - **0.68** - 0.72	0.45 - **0.64** - 0.82	0.52 -**0.73** - 0.82
IOS + DLCO	0.66 - **0.70** - 0.75	0.43 - **0.64** - 0.82	0.55 - **0.73** - 0.84
S + BB + IOS	0.65- **0.68** - 0.72	0.38 - **0.62** - 0.79	0.46 - **0.62** - 0.85
S + BB + DLCO	0.67 - **0.70** - 0.74	0.50 - **0.70** - 0.80	0.45 - **0.68** - 0.82
S + IOS + DLCO	0.67 - **0.71** - 0.77	0.60 - **0.80** - 0.93	0.36 - **0.55** - 0.72
BB + IOS + DLCO	0.65 - **0.69** - 0.73	0.45 - **0.64** - 0.82	0.52 - **0.73** - 0.82
S + BB + IOS + DLCO	0.65 - **0.69** - 0.73	0.50 - **0.80** - 0.90	0.45 - **0.64** - 0.82

Table 4.9: Performance of AR vs. Non-AR (Obstructive) classifier different PFT feature set combinations

4.6.5 Discussion of Individual Classifier Results

Detecting Pulmonary Health

When considering machines in isolation, the spirometer performs the best (median AUC of 0.90), with the body box achieving comparable performance (median AUC of 0.89). Impulse oscillometry had worse sensitivity (0.68, compared to 0.76 for the spirometer). DLCO had worse performance overall (median AUC of 0.77).

No combination of machines performed better than the spirometer in isolation. The only improvement upon the spirometer was achieved in sensitivity by the impulse oscillometer and DLCO; however, the interquartile range was much larger (0.66 - 0.81

PFT Machines(s) Used	AUC	Sensitivity	Specificity
S	0.90 - **0.95** - 0.98	1.00 - **1.00** - 1.00	0.73 - **0.82** - 0.91
BB	0.81 - **0.91** - 0.95	0.88 - **1.00** - 1.00	0.82 - **0.91** - 1.00
IOS	0.89 - **0.95** - 0.98	1.00 - **1.00** - 1.00	0.73 - **0.91** - 1.00
DLCO	0.89 - **0.94** - 0.96	1.00 - **1.00** - 1.00	0.78 - **0.89** - 0.92
S + BB	0.82 - **0.95** - 0.98	1.00 - **1.00** - 1.00	0.82 - **0.91** - 1.00
S + IOS	0.93 - **0.95** - 1.00	1.00 - **1.00** - 1.00	0.73 - **0.86** - 1.00
S + DLCO	0.67 - **0.72** - 0.77	1.00 - **1.00** - 1.00	0.78 - **0.89** - 0.89
BB + IOS	0.77 - **0.91** - 0.98	0.50 - **1.00** - 1.00	0.82 - **0.91** - 1.00
BB + DLCO	0.90 - **0.95** - 1.00	1.00 - **1.00** - 1.00	0.80 - **0.90** - 1.00
IOS + DLCO	0.90 - **0.95** - 1.00	1.00 - **1.00** - 1.00	0.80 - **0.90** - 1.00
S + BB + IOS	0.89 - **0.93** - 1.00	1.00 - **1.00** - 1.00	0.64 - **0.91** - 1.00
S + BB + DLCO	0.89 - **0.94** - 0.94	1.00 - **1.00** - 1.00	0.78 - **0.89** - 0.89
S + IOS + DLCO	0.89 - **0.94** - 1.00	1.00 - **1.00** - 1.00	0.78 - **0.89** - 1.00
BB + IOS + DLCO	0.85 - **0.95** - 1.00	1.00 - **1.00** - 1.00	0.70 - **0.90** - 1.00
S + BB + IOS + DLCO	0.89 - **0.94** - 1.00	1.00 - **1.00** - 1.00	0.78 - **0.89** - 1.00

Table 4.10: Performance of COPD vs. Asthma (Comorbid) classifier different PFT feature set combinations

- 0.88 vs. 0.71 - 0.76 - 0.82).

From the coefficient analysis, there are two clear features that helped detect unhealthy patients: DLCO-SB (from DLCO) and FEV1 (from the spirometer). The top two features for detecting healthy patients were KCO (from DLCO) and FEV1.

Detecting Obstructive Pulmonary Disease

When considering machines in isolation, the spirometer performed significantly better than the other machines (median AUC of 0.94). Like the the pulmonary health classifier, DLCO performed the worst (median AUC of 0.66).

Only the spirometer combined with the impulse oscillometer achieved better per-

PFT Machines(s) Used	AUC	Sensitivity	Specificity
S	0.92 - **0.96** - 1.00	0.83 - **1.00** - 1.00	0.88 - **0.88** - 1.00
BB	0.85 - **0.92** - 0.95	0.83 - **0.83** - 1.00	0.78 - **0.89** - 1.00
IOS	0.85 - **0.90** - 0.94	0.83 - **1.00** - 1.00	0.67 - **0.78** - 0.89
DLCO	0.81 - **0.89** - 0.95	0.80 - **0.80** - 1.00	0.63 - **0.88** - 0.88
S + BB	0.86 - **0.91** - 0.94	0.83 - **0.83** - 1.00	0.78 - **0.89** - 0.92
S + IOS	0.96 - **0.99** - 1.00	1.00 - **1.00** - 1.00	0.88 - **1.00** - 1.00
S + DLCO	0.91 - **0.97** - 1.00	0.75 - **1.00** - 1.00	0.88 - **1.00** - 1.00
BB + IOS	0.91 - **0.94** - 0.97	0.83 - **0.92** - 1.00	0.78 - **0.89** - 0.89
BB + DLCO	0.87 - **0.93** - 0.96	0.75 - **0.75** - 1.00	0.86 - **0.86** - 1.00
IOS + DLCO	0.93 - **0.96** - 1.00	0.75 - **1.00** - 1.00	0.86 - **1.00** - 1.00
S + BB + IOS	0.94 - **0.98** - 1.00	0.80 - **1.00** - 1.00	0.88 - **1.00** - 1.00
S + BB + DLCO	0.88 - **0.97** - 1.00	0.75 - **1.00** - 1.00	1.00 - **1.00** - 1.00
S + IOS + DLCO	0.94 - **0.98** - 1.00	0.75 - **1.00** - 1.00	0.86 - **1.00** - 1.00
BB + IOS + DLCO	0.95 - **0.98** - 1.00	0.75 - **1.00** - 1.00	0.86 - **1.00** - 1.00
S + BB + IOS + DLCO	0.96 - **1.00** - 1.00	0.75 - **1.00** - 1.00	1.00 - **1.00** - 1.00

Table 4.11: Performance of COPD vs. Asthma (Non-comorbid) classifier different PFT feature set combinations

formance than the spirometer in isolation; however, the improvement was minimal (median AUC of 0.95). Overall, just like with the pulmonary health classifier, the best performance was achieved with just the spirometer.

From the coefficient analysis, the most important feature came from the spirometer (MMEF 75/25), which was specifically useful for detecting non-obstructive pulmonary diseases. The top feature for detecting obstructive diseases was DLCO (DLCO-SB).

PFT Machines(s) Used	AUC	Sensitivity	Specificity
S	0.75 - **0.80** - 0.88	0.69 - **1.00** - 1.00	0.60 - **0.80** - 0.85
BB	0.70 - **073** - 0.80	0.50 - **0.75** - 1.00	0.40 - **0.60** - 0.80
IOS	0.75 - **0.79** - 0.88	0.50 - **0.75** - 1.00	0.67 - **0.83** - 0.83
DLCO	0.81 - **0.88** - 0.94	0.50 - **0.75** - 1.00	0.75 - **1.00** - 1.00
S + BB	0.73 - **0.78** - 0.85	0.50 - **0.75** - 1.00	0.60 - **0.80** - 1.00
S + IOS	0.78 - **0.85** - 0.95	0.75 - **1.00** - 1.00	0.60 - **0.80** - 1.00
S + DLCO	0.80 - **0.84** - 0.91	0.50 - **0.75** - 1.00	0.75 - **0.75** - 1.00
BB + IOS	0.78 - **0.88** - 0.91	0.50 - **0.75** - 0.81	0.75 - **1.00** - 1.00
BB + DLCO	0.72 - **0.81** - 0.88	0.50 - **0.75** - 0.81	0.75 - **1.00** - 1.00
IOS + DLCO	0.75 - **0.81** - 0.88	0.50 - **0.75** - 0.81	0.75 - **1.00** - 1.00
S + BB + IOS	0.81 - **0.91** - 0.97	0.75 - **0.75** - 1.00	0.75 - **1.00** - 1.00
S + BB + DLCO	0.75 - **0.81** - 0.88	0.50 - **0.75** - 1.00	0.75 - **0.75** - 1.00
S + IOS + DLCO	0.75 - **0.81** - 0.88	0.50 - **0.75** - 1.00	0.75 - **0.75** - 1.00
BB + IOS + DLCO	0.83 - **0.88** - 0.96	0.75 - **0.75** - 1.00	0.67 - **1.00** - 1.00
S + BB + IOS + DLCO	0.88 - **0.92** - 1.00	0.75 - **1.00** - 1.00	0.67 - **1.00** - 1.00

Table 4.12: Performance of AR vs. Non-AR (Non-obstructive) classifier different PFT feature set combinations

Detecting AR

The detection of allergic rhinitis under the obstructive branch (that is, the detection of allergic rhinitis in patients who also have COPD or asthma) is the bottleneck in overall system performance. When all PFT machines are used, the classifier achieved poor results (median AUC: 0.69). This low performance was slightly ameliorated by using the spirometer in isolation, but even then the performance was moderate at best (median AUC: 0.72). We expect most patients to go through this branch, and its poor performance predicts inaccurate results when the full diagnostic protocol is used in the field.

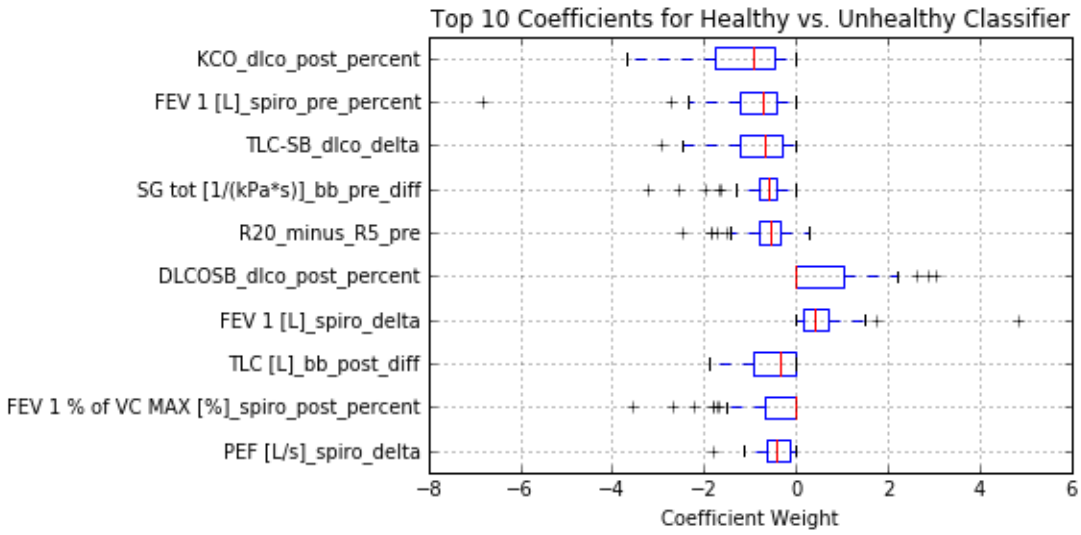

Figure 4-5: Top 10 coefficients of pulmonary health classifier when utilizing features extracted from all PFT machines

The detection of allergic rhinitis under the non-obstructive branch achieves good performance. The best tool in isolation is the DLCO (median AUC: 0.88). However, its sensitivity is moderate (median sensitivity: 0.75). For the best performance, all machines should be used (median AUC: 0.92, median sensitivity: 1.00, median specificity: 1.00).

The top coefficients for the two allergic rhinitis classifiers (Figures 4-7 and 4-10) reinforce the difficult they have at classification: all median coefficients are around zero. This suggests that none of the PFT features aid in detection of AR.

Detecting COPD and Asthma

Both of the COPD/Asthma classifiers (under the AR and non-AR branches) perform well. In isolation under the AR branch, the best machine was the IOS, which achieved the same median AUC as the spirometer (0.95), but had higher specificity (median

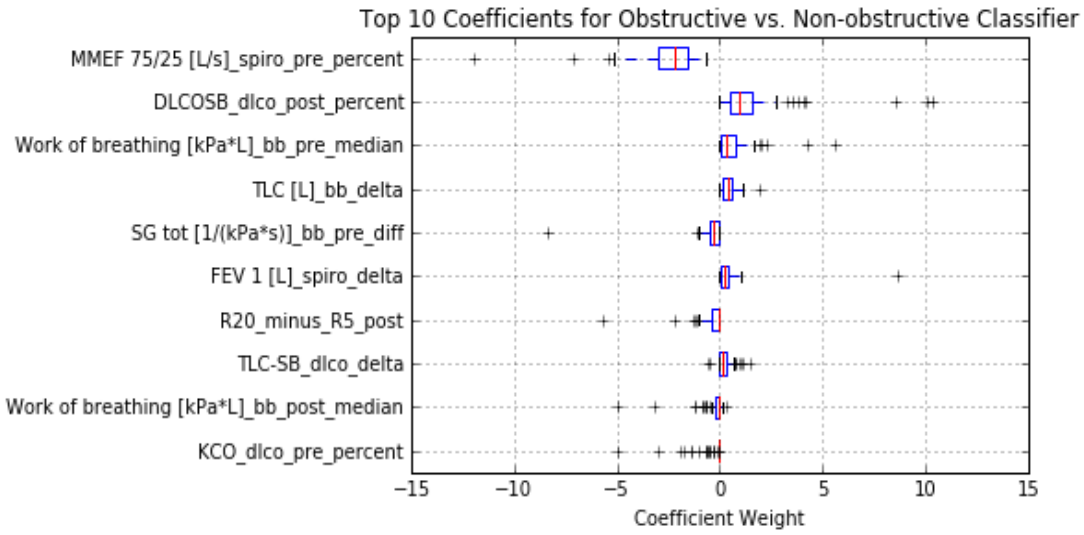

Figure 4-6: Top 10 coefficients of obstructive pulmonary health classifier when utilizing features extracted from all PFT machines

specificty: 0.91). No combination of features achieved better performance than the IOS in isolation.

In isolation under the non-AR branch, the best performance was achieved by the spirometer in isolation (median AUC: 0.96). However, the best performance was achieved by using all the machines (median AUC: 1.00).

The better performance of the COPD/Asthma classifier under the non-AR branch suggests that it is easier to detect patients without comorbidities than those with comorbidities. It reinforces our motivation in our cough analysis to train our classifier on non-comorbid patients.

From the coefficient analysis of the comorbid COPD analysis, TLC-SB (from the DLCO) was the main useful feature. For the non-comorbid analysis, FEV 1 and PEF (both from the spirometer) and RV % TLC and RLC (both from the body box) are the most useful features. Overall, this indicates the importance of the spirometer,

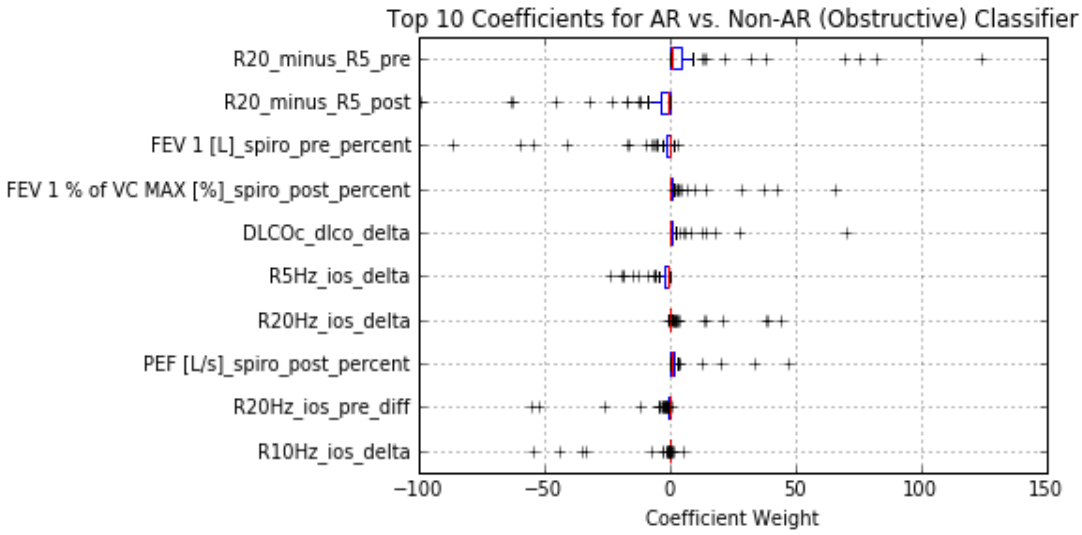

Figure 4-7: Top 10 coefficients of AR vs. Non-AR (Obstructive) classifier when utilizing features extracted from all PFT machines

DLCO, and body box for the detecting of COPD. Specifically, FEV1 and TLC are the most important features.

4.6.6 Recommendations

Diagnostic Value of Individual PFT Machines

For detecting pulmonary health or obstructive pulmonary disease, using more than one machine does not significantly improve performance from the spirometer in isolation. For detecting pulmonary health the spirometer has a median AUC of 0.90; for detecting obstructive pulmonary disease it has a median AUC of 0.94. Overall, as far as these two classifiers are concerned, the spirometer in isolation is sufficient for accurate screening.

Combining more than one machine is useful for the detection of COPD and asthma. If using the fewest number of machines is desired, one should combine the

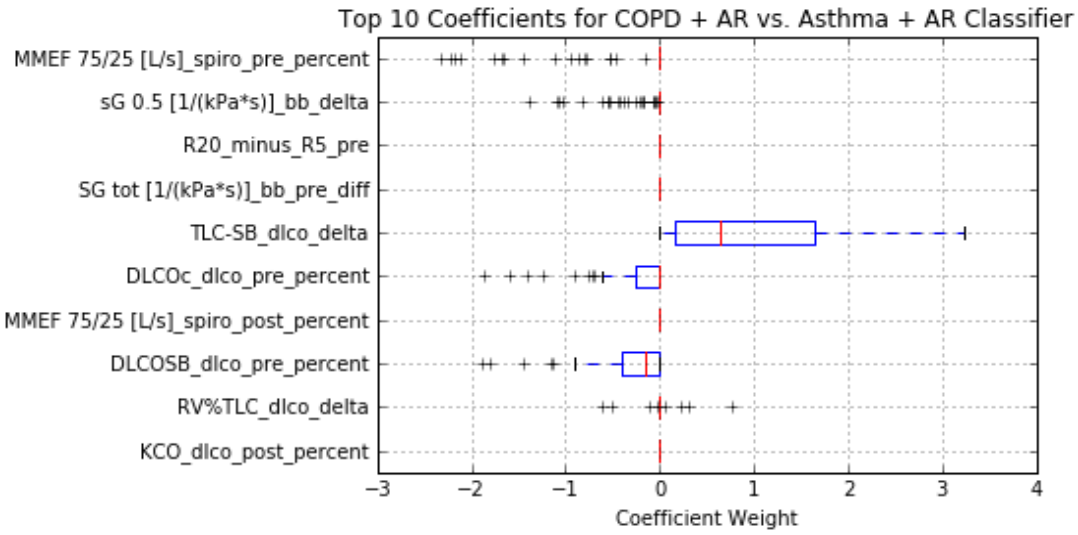

Figure 4-8: Top 10 coefficients of COPD/asthma (comorbid) classifier when utilizing features extracted from all PFT machines

spirometer with either the impulse oscillometer (median AUC of 0.98) or the DLCO (median AUC of 0.97). Practically, the impulse oscillometer should be administered given the brevity, ease, and affordability of the exam relative to the DLCO (which requires a blood exam for hemoglobin calibration).

If the maximum accuracy is desired, then all four machines should be administered – this results in near-perfect classification (median AUC of 0.99).

Recommendation for PFT Feature Selection

From the coefficient analyses for the three classification schemes, multiple features appeared multiple times as the top features. The most useful feature comes from the spirometer (FEV1) – this is an important feature for detecting healthy patients and asthma. DLCO-SB is another important feature, specifically for detecting unhealthy patients and obstructive pulmonary disease. For detecting COPD, TLC is vital

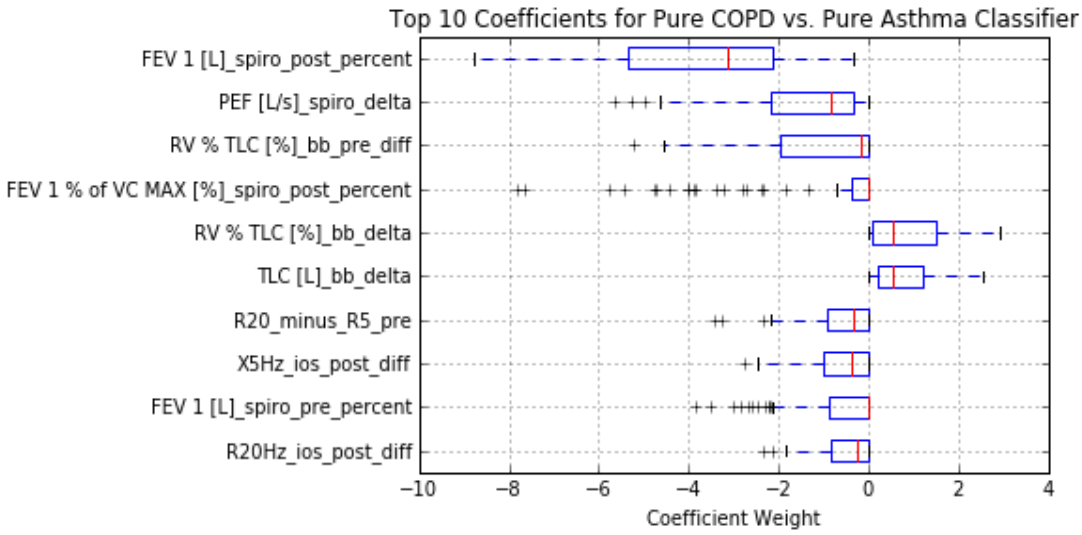

Figure 4-9: Top 10 coefficients of COPD/asthma (pure) classifier when utilizing features extracted from all PFT machines

(either from the DLCO or body box). Finally, the feature MMEF 75/25 from the spirometer is important for detecting non-obstructive pulmonary diseases.

4.7 Supervised Analysis of PFT Data with Clinical Features

4.7.1 Motivation

In this section, we aimed to simulate a more realistic clinical scenario by combining the PFT data with clinical features (questionnaire, lung sounds). We aimed to discover the utility of clinical features for diagnosis in combination with the PFT data.

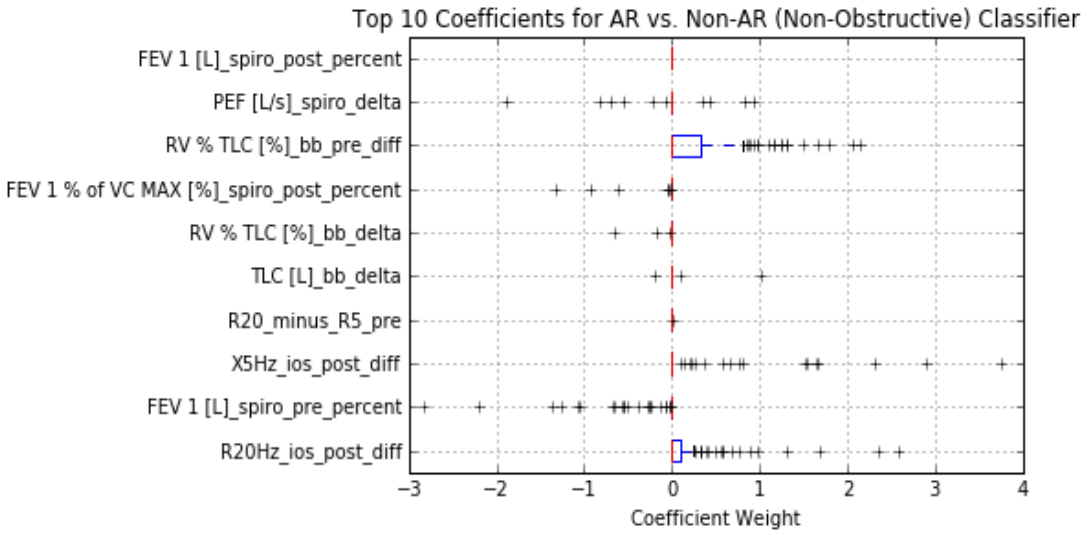

Figure 4-10: Top 10 coefficients of AR vs. Non-AR (Non-obstructive) classifier when utilizing features extracted from all PFT machines

4.7.2 Methods

Similar to the previous section, the following analysis uses the same multi-layer classification shown in 4-4. We analyzed each classifier with the PFT data in isolation, combined with one of either lung sounds or questionnaire data, and then combined with both.

4.7.3 Results of Supervised Learning Analysis

Tables 4.13, 4.14, 4.15, 4.16, 4.17, and 4.18 summarize the performance of the full diagnostic protocol classifiers when trained using various combinations of PFT and clinical data.

Features Used	AUC	Sensitivity	Specificity
PFT	0.88 - **0.90** - 0.93	0.71 - **0.75** - 0.82	0.86 - **0.93** - 1.00
PFT + Q	0.98 - **0.99** - 1.00	0.93 - **0.93** - 0.96	1.00 - **1.00** - 1.00
PFT + L	0.88 - **0.90** - 0.92	0.71 - **0.75** - 0.82	0.86 - **0.93** - 1.00
PFT + L + Q	0.98 - **0.99** - 1.00	0.89 - **0.93** - 0.96	1.00 - **1.00** - 1.00

Table 4.13: Performance of Healthy vs. Unhealthy classifier when trained on different combinations of PFT and clinical data

Features Used	AUC	Sensitivity	Specificity
PFT	0.96 - **0.99** - 1.00	0.86 - **0.93** - 1.00	1.00 - **1.00** - 1.00
PFT + Q	0.98 - **0.99** - 1.00	0.93 - **0.93** - 0.96	1.00 - **1.00** - 1.00
PFT + L	0.97 - **0.99** - 0.99	0.86 - **0.95** - 1.00	1.00 - **1.00** - 1.00
PFT + L + Q	0.99 - **0.99** - 1.00	0.95 - **0.95** - 1.00	1.00 - **1.00** - 1.00

Table 4.14: Performance of Obstructive vs. Non-obstructive classifier when trained on different combinations of PFT and clinical data

4.7.4 Discussion

For the detection of pulmonary health, the questionnaire is critical for good performance. From the cough analysis, we know that the questionnaire in isolation achieves near perfect classification of pulmonary health. For this classifier, lung sounds do not add any benefit to PFT data.

For the detection of obstructive disease, neither lung sounds nor the questionnaire caused an increase in performance.

For the detection of pure COPD and pure asthma, both lung sounds and the questionnaire increase the performance of the PFT data slightly (from a median AUC of 0.98 to 1.00). The clinical features are not required for this classification.

For the detection of allergic rhinitis, both lung sounds and the questionnaire increase performance, though the questionnaire increases the median AUC more (by

Features Used	AUC	Sensitivity	Specificity
PFT	0.66 - **0.70** - 0.74	0.50 - **0.60** - 0.90	0.45 - **0.64** - 0.82
PFT + Q	0.87 - **0.92** - 0.96	0.90 - **1.00** - 1.00	0.73 - **0.82** - 0.91
PFT + L	0.67 - **0.70** - 0.74	0.50 - **0.70** - 0.90	0.45 - **0.89** - 1.00
PFT + L + Q	0.87 - **0.91** - 0.96	0.90 - **0.95** - 1.00	0.70 - **0.82** - 0.90

Table 4.15: Performance of AR classifier (obstructive) when trained on different combinations of PFT and clinical data

Features Used	AUC	Sensitivity	Specificity
PFT	0.89 - **0.94** - 1.00	1.00 - **1.00** - 1.00	0.78 - **0.89** - 1.00
PFT + Q	0.89 - **0.94** - 1.00	1.00 - **1.00** - 1.00	0.78 - **0.89** - 1.00
PFT + L	0.89 - **0.94** - 1.00	1.00 - **1.00** - 1.00	0.78 - **0.89** - 1.00
PFT + L + Q	0.89 - **0.94** - 1.00	1.00 - **1.00** - 1.00	0.78 - **0.89** - 1.00

Table 4.16: Performance of COPD vs. Asthma (comorbid) classifier when trained on different combinations of PFT and clinical data

0.06, instead of 0.02). This indicates that the questionnaire is an important tool for classifying AR.

For the detection of patients with both COPD and AR, lung sounds and the questionnaire do not offer increased performance. The PFT data are ideal for the detection of COPD. However, for the detection of patients with both asthma and AR, the addition of the questionnaire significantly increases performance (the median AUC increases by 0.22).

These results suggest that the addition of clinical features is important to allow the PFT data to accurately detect pulmonary health and allergic rhinitis. The questionnaire, specifically, is the more important tool, and it should be included in any future classifiers trained on PFT data in order to ensure maximum performance.

Features Used	AUC	Sensitivity	Specificity
PFT	0.91 - **0.98** - 1.00	0.75 - **1.00** - 1.00	1.00 - **1.00** - 1.00
PFT + Q	0.98 - **0.99** - 1.00	1.00 - **1.00** - 1.00	1.00 - **1.00** - 1.00
PFT + L	0.96 - **1.00** - 1.00	0.75 - **1.00** - 1.00	0.75 - **1.00** - 1.00
PFT + L + Q	1.00 - **1.00** - 1.00	1.00 - **1.00** - 1.00	1.00 - **1.00** - 1.00

Table 4.17: Performance of COPD vs. Asthma (non-comorbid) classifier when trained on different combinations of PFT and clinical data

Features Used	AUC	Sensitivity	Specificity
PFT	0.83 - **0.94** - 1.00	0.75 - **1.00** - 1.00	0.67 - **1.00** - 1.00
PFT + Q	1.00 - **1.00** - 1.00	1.00 - **1.00** - 1.00	1.00 - **1.00** - 1.00
PFT + L	0.86 - **0.96** - 1.00	0.75 - **1.00** - 1.00	0.67 - **1.00** - 1.00
PFT + L + Q	1.00 - **1.00** - 1.00	1.00 - **1.00** - 1.00	**1.00** - 1.00 - 1.00

Table 4.18: Performance of AR (non-obstructive) classifier when trained on different combinations of PFT and clinical data

4.8 Supervised Learning Analysis of PFT Data Using Independent Classifiers

4.8.1 Motivation

While the full, tree-like diagnostic protocol is medically intuitive, its tree structure exacerbates errors that originate in higher nodes. For example, we have shown that the PFT tools alone are not adequate for the detection of pulmonary health, especially when compared to the risk factor and symptom questionnaire. Since this is the root node of the diagnostic tree, the reduced performance leads to poor accuracy. Therefore, we analyzed disease-specific classifiers trained and tested on all patients as a means of exploring potential improvement in diagnostic prediction.

4.8.2 Classification Design

Unlike the previous supervised learning analyses that used a multi-layer classification structure, in this analysis we explored a single-layer classifier for each disease. Although this approach is less informative to the clinician, it eliminates the compounding of errors produced by the multi-layer design.

4.8.3 Methods

We trained three binary classifiers for the detection of COPD, asthma, and allergic rhinitis, using all Mobile Kit tools, and compared the results when trained on all PFT data. The classifiers trained on the Mobile Kit were trained and tested using the entire dataset. The classifiers trained on the entire PFT dataset, but tested using only the patients which had a complete set of PFT data, Mobile Kit data, and cough sound data.

In addition, we also explored the effect of comorbidities on the accuracy of the classifiers. We did so by checking the union of the COPD and AR classifiers and the union of the asthma and AR classifiers. For these patients, we can only report the sensitivity.

4.8.4 Results

4.8.5 Discussion

The full diagnostic protocol contains two bottlenecks in performance: the root node which detects pulmonary health, and the AR classifier under the obstructive branch.

In the multi-layer classifier structure, the relatively poor performance of these classifiers degrades the overall performance of all the subsequent layers underneath.

Classifier	Sensitivity (PFT)	Specificity (PFT)
Unhealthy	0.88	1.00
AR	0.71	0.83
Asthma	0.93	1.00
COPD	1.00	1.00
Asthma + AR	0.73	N/A
COPD + AR	0.63	N/A

Table 4.19: Performance (sensitivity and specificity) of independent classifiers when trained using data from all PFT machines

The AR classifier increases slightly in performance when diagnosed with an independent classifier (Table 4.19). Nonetheless, the AR classifier still only achieves moderate accuracy.

Unlike the AR classifier, the COPD and asthma classifiers do achieve good performance, especially the one for COPD.

The performance of comorbid detection (AR plus either COPD or asthma) is moderate, given the inability of the PFT data to properly detect AR.

Overall, these results suggest that having multiple independent classifiers might lead to better overall performance, as opposed to the tree-like diagnostic protocol. While it may not be the preferred setup from a medical perspective, it provides an easy and effective solution to the lowered performance of the full protocol.

4.9 Performance Comparison of PFT vs. Mobile Kit

In order to analyze the entire classification scheme, we trained the scheme's classifiers using all of the PFT data and simulated the classification on a subset of patients (meaning that if a patient was misclassified in a higher node, the patient did not

go through the classifiers of lower nodes). We repeated the analysis using classifiers trained on data from all of the Mobile Kit's tools.

4.9.1 Multi-Layer Classifier Structure

Table 4.20 shows the confusion matrix from the simulation run using the multi-layer classifiers trained on the PFT data, while Table 4.21 shows the confusion matrix from the simulation run using the classifiers trained on the Mobile Kit data.

Compared to the PFT tools in isolation, the Mobile Kit has a clear advantage at detecting pulmonary health. Even though the supervised analysis showed that the classifiers trained on the PFT data would perform better at the disease-level, the inability of these classifiers to accurately detect pulmonary health prematurely removes patients from further analysis and leads to poor performance.

When compared to the PFT integrated with clinical features, the Mobile Kit performs relatively well, but has degraded performance in the ability to distinguish between specific diseases.

This indicates that the independent classifier approach might be preferable to the tree-structure of diagnosis, since it removes the interdependencies of the classifiers.

4.9.2 Independent Classifiers

Table 4.22 shows the confusion matrix from the simulation run using independent classifiers trained on the PFT data compared to when they were trained on the Mobile Kit data.

In summary, we see that while the Mobile Kit has better performance for classifying Unhealthy and AR patients, the PFT exhibits slightly better performance for distinguishing between asthma and COPD.

		Predicted						
		Healthy	AR	Asthma	COPD	Asthma + AR	COPD + AR	Other
Actual	Healthy	9	1	1	3	2	1	2
	AR	2	0	0	1	1	0	1
	Asthma	0	0	4	6	4	5	2
	COPD	0	0	3	4	1	4	0
	Asthma+AR	0	0	4	9	6	6	3
	COPD+AR	0	0	0	1	0	1	0
	Other	2	0	0	1	1	1	0

Table 4.20: Classification results of running patients through full diagnostic protocol trained on PFT data

We see that the use of independent classifiers reduces the misclassification error overall. The implications of these results for the design of the Mobile Diagnostic Kit is discussed in the next chapter.

		Predicted						
		Healthy	AR	Asthma	COPD	Asthma + AR	COPD + AR	Other
Actual	Healthy	15	0	0	0	0	0	0
	AR	0	2	0	2	2	1	3
	Asthma	0	0	6	7	6	5	3
	COPD	0	0	4	3	2	4	2
	Asthma+AR	0	0	1	10	8	2	10
	COPD+AR	0	0	1	2	1	1	2
	Other	2	0	4	3	4	3	1

Table 4.21: Classification results of running patients through full diagnostic protocol trained on Mobile Kit data

Classifier	**Sensitivity (PFT)**	**Specificity (PFT)**	**Sensitivity (Mobile Kit)**	**Specificity (Mobile Kit)**
Unhealthy	0.88	1.00	0.93	1.00
AR	0.71	0.83	0.80	1.00
Asthma	0.93	1.00	0.89	0.87
COPD	1.00	1.00	1.00	0.87
Asthma + AR	0.73	N/A	0.77	N/A
COPD + AR	0.63	N/A	0.40	N/A

Table 4.22: Performance (sensitivity and specificity) of independent classifiers when trained using PFT data vs. Mobile Kit tools

Chapter 5

Summary of Findings and Recommendation for Mobile Diagnostic Kit

5.1 Summary of Findings

The work of this thesis had three main contributions: the development of an allergic rhinitis classifier, the exploration of cough sounds as an additional feature for pulmonary disease diagnosis, and the analysis of pulmonary function testing (PFT) for further validation of the Mobile Kit. These topics are briefly summarized below.

5.1.1 Allergic Rhinitis Classifier

An allergic rhinitis classifier has been analyzed both in isolation and integrated in the full diagnostic protocol. While the integrated classifier creates a bottleneck in overall system performance, the isolated classifier performs well.

5.1.2 Cough Sound Analysis

Cough sounds have been analyzed as a new source for the Mobile Kit. In isolation, cough sound analysis provides moderate performance for pulmonary disease diagnosis. Except in addition to lung sounds, cough sounds do not have any added benefit with the other Mobile Kit tools. However, our analysis of cough sounds has identified an opportunity to create a simplified diagnostic kit that can be used by low-skilled community health care workers.

5.1.3 Pulmonary Function Testing (PFT) Analysis

Pulmonary function testing (PFT) data have been used to train the same classifiers as the Mobile Kit. Our results showed that the Mobile Diagnostic Kit developed in our group performs well compared to the gold standard PFT lab, with slightly degraded performance in distinguishing between specific disease categories.

A more significant implication of our analysis revolves around the use of a multi-layer classifier instead of a single-layer classifier. While the multi-layer classifier provides helpful diagnostic guidance to the clinician, it suffers from lowered performance compared to the individual single-layer classifiers for each disease. This is a general observation of many machine learning systems which will need to be further discussed for the design of future diagnostic systems.

5.2 Mobile Pulmonary Screening Application

5.2.1 Current Application Design

The Android implementation of the Pulmonary Screener was developed by Daniel Chamberlain. It involves a user interface which requests the information of the Mobile Kit (questionnaire, peak flow meter, lung sounds), with the machine learning models working in the background.

The application works with the old classification scheme, summarized in Figure 5-1. There is a script which converts the coefficient weights and intercept of each logistic regression model into XML, a format readable by Android code.

For each model, there is an ArrayList of features (such as questionnaire questions) ordered from largest magnitude of weight to smallest. In essence, for each model, the application asks for features in order of importance. This is to make the process as quick as possible, for once the application is certain enough of its classification (defined as crossing a 50% threshold), it moves on to the next model if the classification is positive (negative classifications, under the old scheme, will never appear in lower leaves of the diagnostic tree).

The Android code is modeled as per Figure 5-2, taken from Daniel Chamberlain's thesis.

5.2.2 Recommendations for Future Versions

Allergic Rhinitis Classification

From the work of this thesis, there are three possible directions that a future version of the Pulmonary Screener can take with regards to allergic rhinitis classification. They are summarized below.

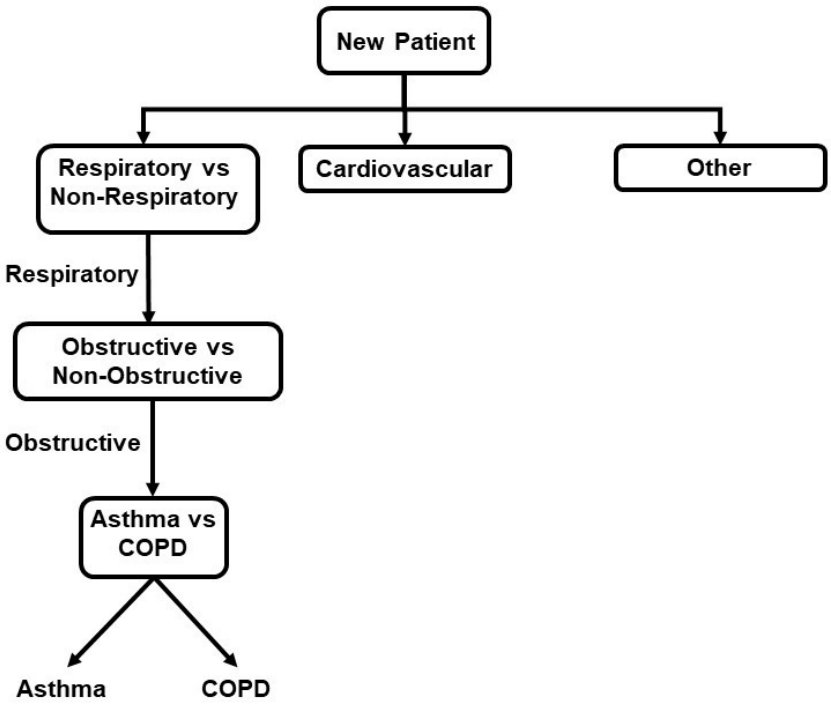

Figure 5-1: Previous classification scheme.

Scenario #1: Integrate Allergic Rhinitis Classifier into Classification Scheme

The first scenario is to integrate the AR classifier within the diagnostic tree, as summarized in Figure 5-3. Six models will have to be saved within the application as XML files, parsed by ModelParser and used by the Model object.

Additional logic will have to be added to determine when to call subsequent models. At the moment, the application proceeds to the next model only if the current model has a positive result. However, under this scenario, the obstructive disease model would have to proceed to one of two AR classifiers depending on whether the

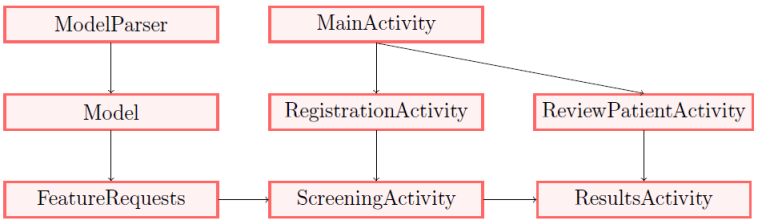

Figure 5-2: Figure from Daniel Chamberlain's thesis summarizing the workflow of the Pulmonary Screener Application.

application determines a patient as having an obstructive or non-obstructive disease. Similarly, the AR classifier under the obstructive branch needs to proceed to one of two COPD/Asthma classifiers depending on whether the patient is classified as either AR or not. This can be done via simple conditional statements.

While this scenario is intuitive for physicians, the performance of the AR classifier under the obstructive branch is moderate at best. In order to ensure the optimum performance of the Android application, this scenario is not recommended until the performance of this classifier can be improved.

Scenario #2: Make All Classifiers Independent

The second scenario is to replace the diagnostic tree with independent and disease-specific classifiers. There would be four models saved within the application as XML files: pulmonary health, COPD, asthma, and allergic rhinitis.

Give that there would no longer be any dependence among the classifiers, the user can decide which classifier(s) to run. This can be achieved by having multiple buttons that call up the appropriate model and run the algorithm. The question of comorbidities can be handled in this case by simply calling both the AR classifier and either the COPD or asthma classifier. In order to minimize the number of questions asked, the same protocol used in the current model (of asking questions in order of

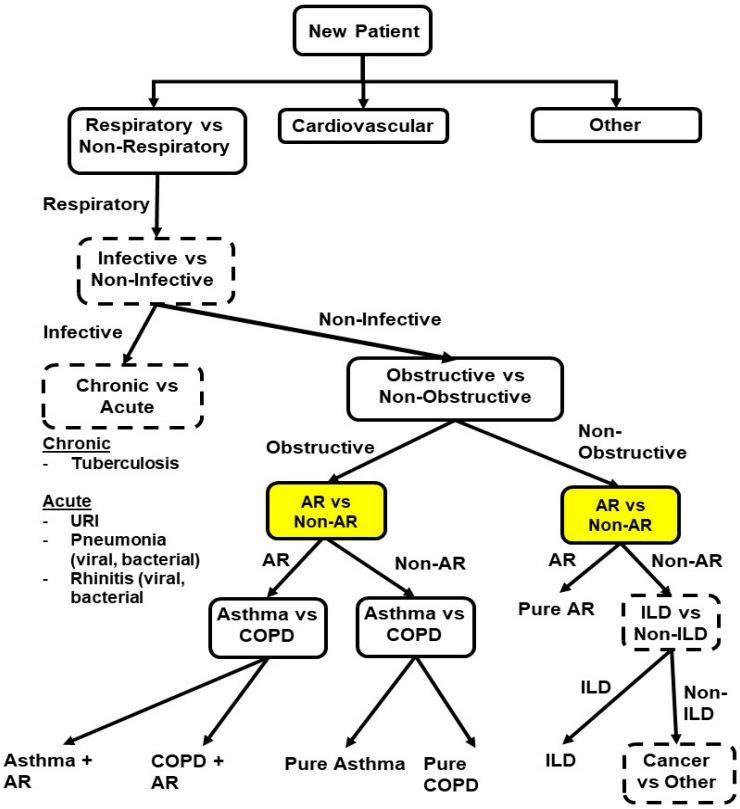

Figure 5-3: New classification scheme. Dotted lines denote classifiers that will be created in the future. Yellow boxes denote additions to the old scheme.

importance and stopping once a 50% threshold is passed) would be used.

While this scenario would provide the optimum performance for disease diagnosis, it is not intuitive for physicians. However, we are still able to provide the most relevant coefficients used for determining the classification. This scenario would not be a black box; physicians will be given feedback as to the algorithm's logic, providing a sense of interpretability to the application.

Scenario #3: Add Independent Allergic Rhinitis Classifier

The third scenario is a hybrid of the two scenarios above. It would leave the application as it exists, and add an extra model (for AR diagnosis) that is always run. One can imagine adding an initial model which diagnoses AR and then, regardless of outcome, begins the Pulmonary Health classifier and runs through the rest of the current application's logic as normal. In the end, the application will provide both the outcome of the AR classifier and the current diagnostic protocol.

Presently, his is the recommended approach. It avoids the performance bottleneck of integrating the AR classifiers within the diagnostic tree while also avoiding the cumbersome nature of having to manually run individual classifiers. As far as the use is concerned, there would be no change to the application; all changes would occur in the back-end.

Addition of Cough Sound Analysis

While cough sounds were not found to have an added benefit to the Mobile Diagnostic Kit, they did achieve moderate performance in isolation. As mentioned previously, we have identified an the potential for an alternative Mobile Diagnostic Kit based on cough sounds alone, which would be easier and quicker to use than a full questionnaire, peak flow meter, and auscultation. Such a kit could be used by low-skilled community health care workers. In terms of the software implementation, this can be done via the same platform which the current application currently uses to collect the peak flow meter and lung sound data. While we do not expect the performance of this simplified tool to be equivalent to the full diagnostic kit, it is an option for areas where using some of the Mobile Kit's tools is impractical (for example, administering the questionnaire to children, buying the electronic stethoscope,

or undergoing the peak flow meter tests).

5.3 Summary

In this thesis, we have presented several new contributions which can be integrated into our current Mobile Diagnostic Kit for pulmonary disease. These new contributions include: 1) expanding our diagnosis classification to include allergic rhinitis and comorbidities, in addition to asthma and COPD; 2) the use of cough sounds as a potential simplified pulmonary disease screening tool for use by community health care workers; 3) validation of our Mobile Diagnostic Kit against the gold standard pulmonary function testing (PFT) lab; and 4) exploration of the use of single-layer, disease-specific classifiers to be used in conjunction with the current multi-layer classification design. Going forward, we feel that these contributions will play an important role in improving the performance and utility of our Mobile Diagnostic Kit.

Bibliography

[1] Thorpe W, Kurver M, King G, Salome C. *Acoustic analysis of cough.* Intelligent Information Systems Conference, The Seventh Australian and New Zealand 2001; 2001. p. 391-394

[2] J. Korpas, J. Sadlonova, M. Vrabec. *Analysis of the cough sound: an overview.* Pulmonary Pharmacology, 1996.

[3] World Health Organization (WHO). *Chronic respiratory diseases: burden of COPD.* WHO. [Online]. Available: http://www.who.int/respiratory/copd/burden/en/

[4] PR Newswire *India MOTECH* http://www.prnewswire.com/news-releases/nationwide-launch-of-mobile-health-program-in-rural-india-signals-new-era-of-mhealth-for-emerging-economies-300204975.html

[5] World Health Organization (WHO) *WHO / Asthma* http://www.who.int/respiratory/asthma/en/

[6] World Health Organization (WHO) *WHO / COPD* http://www.who.int/respiratory/copd/burden/en/

[7] European Respiratory Society (ERS) *The burden of lung disease* http://www.erswhitebook.org/chapters/the-burden-of-lung-disease/

[8] National Heart, Lung, and Blood Institute *COPD* https://www.nhlbi.nih.gov/health/health-topics/topics/copd

[9] S. Braman. *The global burden of asthma.* Chest, Jul. 2016

[10] World Health Organization (WHO), Global Surveillance. *Prevention and Control of Chronic Respiratory Diseases: A Comprehensive Approach.* World Health Organization, 2007.

[11] World Health Organization (WHO). *WHO | The top 10 causes of death . WHO. [Online]. Available: http://www.who.int/mediacentre/factsheets/fs310/en/.*

[12] European Lung Society (ERS). *The burden of lung disease*. [Online]. Available: http://www.erswhitebook.org/chapters/the-burden-of-lung-disease/)/

[13] American Academy of Allergy Asthma and Immunology *COPD*. https://www.nhlbi.nih.gov/health/health-topics/topics/copd

[14] Grameen Foundation. *Nationwide launch of mobile health program in rural India signals new era of mHealth for emerging economies*. Available: http://www.prnewswire.com/news-releases/nationwide-launch-of-mobile-health-program-in-rural-india-signals-new-era-of-mhealth-for-emerging-economies-300204975.html

[15] L. Abdoune, M. Fezari. *A sound database for health smart home*. Proceedings of the World Congress on Computer Applications and Information Systems (WCCAIS), Jan. 2014

[16] Y. Amrulloh, U. Abeyratne, V. Swarnkar, R. Triasih, A. Setyah. *Automatic cough segmentation from non-contact sound recordings in pediatri wards*. Biomedical Signal Processing and Control, Aug. 2015

[17] C. Infante, D. Chamberlain, R. Kodgule, R. Fletcher. *Classification of voluntary coughs applied to the screening of respiratory disease*. IEEE EMBC, 2017

[18] Y. Amrulloh, U. Abeyratne, V. Swarnkar, R. Triasih. *Cough sound analysis for pneumonia and asthma classification in pediatric population*. 6th International conference on Intelligent Systems, Modelling and Simulation, 2015

[19] R. Pramono, S. Imtiaz, E. Rodriguez-Villegas. *A Cough-Based Algorithm for Automatic Diagnosis of Pertussis*. PLOS One, Sept. 2016

[20] D. Chamberlain, R. Kodgule, R. Fletcher. *A mobile platform for automated screening of asthma and chronic obstructive pulmonary disease*. IEEE EMBC, 2016

[21] A. Abaza, J. Day, J. Reynolds, A. Mahmoud, W. Goldsmith, W. McKinney, E. Petsonk, D. Frazer. *Classification of voluntary cough sound and airflow patterns for detecting abnormal pulmonary function*. Cough, Nov. 2009

[22] S. Aydore, I. Sen, Y. P. Kahya, and M. K. Mihcak. *Classification of respiratory signals by linear analysis.* Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2009. EMBC 2009, 2009, pp. 2617âĂŞ2620

[23] C. Infante, D. Chamberlain, R. Kodgule, R. Fletcher. *Classification of voluntary coughs applied to the screening of respiratory disease.* IEEE EMBC, 2017

[24] V. Swarnkar, U. Abeyratne, Y. Amrulloh, C. hukins, R. Triasih, A. Setyati. *Neural network based algorithm for automatic identification of cough sounds.* 35th Annual International Conference of the IEEE EMBS, July 2013.

[25] [20] R. Pellegrino, G. Viegi, V. Brusasco, R. O. Crapo, F. Burgos, R. Casaburi, A. Coates, C. P. M. v. d. Grinten, P. Gustafsson, J. Hankinson, R. Jensen, D. C. Johnson, N. MacIntyre, R. McKay, M. R. Miller, D. Navajas, O. F. Pedersen, and J. Wanger. *Interpretative strategies for lung function tests* European Respiratory Journal, vol. 26, pp. 948âĂŞ968, Nov. 2005.

[26] J. Smith, H. Ashurst, S. Jack, A. Woodcock, and J. Earis. *The description of cough sounds by healthcare.*

[27] S. Abeyratne, Y. Amrulloh, A. Chang *Automated algorithm for Wet/Dry cough sounds classification.* 34th Annual International Conference of the IEEE EMBS, 2012.

[28] H. Chatrzarrin, A. Arcelus, R. Goubran, F. Knoefel *Feature Extraction for the Differentiation of Dry and Wet Cough Sounds.* Medical Measurements and Applications Proceedings (MeMeA), 2011.

[29] Y. Amrulloh, D. Wati, F. Pratiwi, R. Triasih *A Novel Method for Wet/Dry Cough Classification in Pediatric Population.* IEEE Region 10 Symposium (TENSYMP), 2016.

[30] D. Chamberlain, R. Kodgule, and R. Fletcher. *Towards a Pulmonary Diagnostic Kit for Telemedicine and Global Health Point-Of-Care Diagnosis.* NIH-IEEE 2015 Strategic Conference on Healthcare Innovations and Point-of-Care Technologies for Precision Medicine, 2015

[31] D. Chamberlain, R. Kodgule, D. Ganelin, V. Miglani, and R. Fletcher. *Application of Semi-Supervised Deep Learning to Lung Sound Analysis.* Engineering in Medicine and Biology Society (EMBC), IEEE 38th Annual International Conference, 2016.

[32] Hanieh Chatrzarrin, Amaya Arcelus, Rafik Goubran, Frank Knoefel. *Feature Extraction for the Differentiation of Dry and Wet Cough Sounds*. NSERC IEEE, pp. 162âĂŞ166, Aug 2011.

[33] V. Singh, V. Mittal. *Preliminary Analysis of Cough Sounds*. IEEE INDICON, 2015.

[34] *Management of Allergic Rhinitis and Its Impact on Asthma: Pocket Guide*. Allergic Rhinitis and Its Impact on Asthma, 2008.

[35] A. Ghosal, GD Ravindran, P. Gangwal, G. Rajadhyaksha, S. Cho, A. Muttalif, H. Lin, S. Thanaviratananich, S. Bagga, R. Faruqi, S. Sajjan, P. Shetty, R. Syed, K. Hamrosi, D. Wang. *The burden of segregated respiratory diseases in India and the quality of care in these patients: Results from the Asia-Pacific Burden of Respiratory Diseases study*. Lung India (Indian Chest Society), Vol. 33, Issue 6. 2016.

[36] Burke, Jason. *India's Doctors Blame Air Pollution for Sharp Rise in Respiratory Diseases*. The Guardian. Guardian News and Media, 23 Sept. 2015.

[37] C. Chauhan. *13 out of World's Top 20 Polluted Cities in India, Only Three in China*. Hindu Times. 04 June 2015.

[38] Hu G, Zhou Y, Tian J, Yao W, Li J, Li B, Ran P Chest. 2010 Jul; 138(1):20-31.

[39] Prasad R, Singh A, Garg R, Giridhar GB. Biosci Trends. 2012 Oct; 6(5):219-28.

[40] Mumbai: IIPS *International Institute of Population Sciences (IIPS) and Macro International*. National Family Health Survey (NFHS-3, 2005-2006.: India)

[41] Koul PA, Hajni MR, Sheikh MA, Khan UH, Shah A, Khan Y, Ahangar AG, Tasleem RA. *Hookah smoking and lung cancer in the Kashmir valley of the Indian subcontinent.*. Asian Pac J Cancer Prev. 2011; 12(2):519-24.

[42] ICMR-MRC Workshop. *Building Indo-UK Collaboration in chronic diseases*. 2009

[43] A. Bhome. *COPD in India: Iceberg or volcano?*. Journal of Thoracic Disease. Vol. 4, No. 3. 2012.

[44] Lopez AD, Shibuya K, Rao C, Mathers CD, Hansell AL, Held LS, et al. *Chronic obstructive airway disease: Current burden and future projections*. Eur Resp J. 2006;27:397âĂŞ412.

[45] R. Pellegrino, G. Viegi, V. Brusasco, R. O. Crapo, F. Burgos, R. Casaburi, A. Coates, C. P. M. v. d. Grinten, P. Gustafsson, J. Hankinson, R. Jensen, D. C. Johnson, N. MacIntyre, R. McKay, M. R. Miller, D. Navajas, O. F. Pedersen, and J. Wanger. *Interpretative strategies for lung function tests.* European Respiratory Journal, vol. 26, pp. 948âĂŞ968, Nov. 2005.

[46] D. Sadler. *'Cough into this phone, please': Perth startup's diagnosis app could change the way respiratory problems are detected.* Startup Smart.

[47] *Symptom-based integrated approach to the adult in primary care.* https://www.health-e.org.za/wp-content/uploads/2015/04/PC-101-Guideline-v2-2013-14-2.pdf. 2013

[48] *Symptom-based integrated approach to the adult in primary care.* https://www.health-e.org.za/wp-content/uploads/2015/04/PC-101-Guideline-v2-2013-14-2.pdf. 2013.

[49] P. Koul. *Chronic obstructive pulmonary disease: Indian guidelines and the road ahead.* Lung India. July 2013.

[50] *COPD-6.* https://vitalograph.ie/product/161419/copd-6

[51] C. Represas, A. Fernandez-Villar, A. Ruano-Ravina, A. Priegue-Carrera, M. Botana-Rial. *Screening for Chronic Obstructive Pulmonary Disease: Validity and Reliability of a Portable Device in Non-Specialized Healthcare Setting* PLOS ONE, 2016.

[52] Yawn BP, Mapel DW, Mannino DM, Martinez FJ, Donohue JF, Hanania NA, Kosinski M, Rendas-Baum R, Mintz M, Samuels S, Dalal AA. *Development of the lung function questionnaire (LFQ) to identify airflow obstruction.* Int J Chron Obstruct Pulmon Dis. 2010;5:1-10.

[53] Martinez FJ, Raczek AE, Seifer FD, Conoscenti CS, Curtice TG, D'Eletto T, Cote C, Hawkins C, Phillips AL. *Development and initial validation of a self-scored COPD Population Screener Questionnaire (COPD-PS).* COPD. 2008;5(2):85-95.

[54] G. Ortiz. *Tools for Diagnosis and Assessment of Asthma and Chronic Obstructive Pulmonary Disease in Clinical Practice.* Focus on Asthma and COPD. Issue 4

[55] R. Irwin et al. *Diagnosis and Management of Cough.* Chest. Jan. 2006.

[56] *EKO Devices - DUO*. https://ekodevices.com/duo/

[57] *StethoMe*. http://stethome.com/

[58] World Health Organization (WHO) *Density per 1000 Data by country* http://apps.who.int/gho/data/node.main.A1444

[59] M. Topalovic, J. Aerts, T. Troosters, M. Decramer, W. Janssens. *Artificial intelligence to improve the diagnostic power of complete pulmonary function tests* European Respiratory Journal, 2016.

[60] R. Karakis, I. Guler, A. Isik. *Feature selection in pulmonary function test data with machine learning methods* Signal Processing and Communications Applications Conference (SIU), 2013.