

A Mobile Platform for Automated Screening of Asthma and Chronic Obstructive Pulmonary Disease

Daniel B. Chamberlain, Dr. Rahul Kodgule, and Richard Ribón Fletcher *Member, IEEE*

Abstract— Chronic Obstructive Pulmonary Disease (COPD) and asthma each represent a large proportion of the global disease burden; COPD is the third leading cause of death worldwide and asthma is one of the most prevalent chronic diseases, afflicting over 300 million people. Much of this burden is concentrated in the developing world, where patients lack access to physicians trained in the diagnosis of pulmonary disease. As a result, these patients experience high rates of underdiagnosis and misdiagnosis. To address this need, we present a mobile platform capable of screening for Asthma and COPD. Our solution is based on a mobile smart phone and consists of an electronic stethoscope, a peak flow meter application, and a patient questionnaire. This data is combined with a machine learning algorithm to identify patients with asthma and COPD. To test and validate the design, we collected data from 119 healthy and sick participants using our custom mobile application and ran the analysis on a PC computer. For comparison, all subjects were examined by an experienced pulmonologist using a full pulmonary testing laboratory. Employing a two-stage logistic regression model, our algorithms were first able to identify patients with either asthma or COPD from the general population, yielding an ROC curve with an AUC of 0.95. Then, after identifying these patients, our algorithm was able to distinguish between patients with asthma and patients with COPD, yielding an ROC curve with AUC of 0.97. This work represents an important milestone towards creating a self-contained mobile phone-based platform that can be used for screening and diagnosis of pulmonary disease in many parts of the world.

I. INTRODUCTION

A. Pulmonary Diseases

Pulmonary diseases are an increasing global health burden and are responsible for more than 14% of deaths worldwide [1]. More than a billion individuals suffer from chronic pulmonary diseases, which, in addition to causing premature death, reduce the quality of life for those afflicted [2]. The burden of these diseases falls disproportionately on individuals in developing countries due to the prevalence of many associated risk factors (e.g. air pollution, biomass cooking stoves, cigarette smoking).

Chronic Obstructive Pulmonary Disorder (COPD) and asthma are the most common pulmonary diseases among adults. COPD is the third leading cause of death worldwide



Figure 1: Mobile stethoscope being used for auscultation.

and chronic asthma afflicts over 300 million people. Thus, as a goal for a mobile screening or automated diagnostic tools, COPD and asthma represent important priorities.

B. The Challenge of Pulmonary Disease Diagnosis

The current gold standard for pulmonary disease diagnostic equipment is a Pulmonary Function Testing (PFT) lab, which includes an array of testing machines used to measure all aspects of lung physiology. These machines include a body plethysmograph, spirometer, impulse oscillometer, and pulmonary gas diffusion. These testing labs are relatively expensive and require specially-trained technicians for operation [3]. In rural areas and developing countries, PFT labs are quite rare; for example, in India, there are fewer than 100 PFT labs serving the entire country of 1.2 billion people.

For a large fraction of the world's population, pulmonary disease is still diagnosed using only a stethoscope. While this practice is effective when a patient is examined by a trained pulmonologist, such specialists are also rare in rural areas in developing countries. Few general practitioner doctors in these areas are trained to diagnose pulmonary disease and many are not even trained in Western medicine. As a result, despite a large prevalence of pulmonary diseases, these illnesses are poorly understood, and frequently misdiagnosed and underdiagnosed [4], [5].

To improve the accuracy of diagnosis, symptom-based guidelines have been developed. Primary Care 101 is a step-by-step guide to the diagnosis of diseases often encountered by primary care physicians and health workers in low-resource settings [6]. It relies on symptoms and easily measureable physiological parameters to arrive at prospective diagnoses. However, it does not capture any

*Research supported by MIT Tata Center, MIT MISTI India Program, Vodafone Americas Foundation, and NIH.

D. Chamberlain is with the Massachusetts Institute of Technology, Cambridge, MA 02139 USA (301-775-6428; e-mail: dbc@mit.edu).

R. Kodgule is with the Chest Research Foundation, Pune, 411014 India (e-mail: drrahul@crfindia.com).

R. Fletcher is with the Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: fletcher@media.mit.edu).

information about abnormal lung sounds or reduced lung function, which are essential for confirming diagnosis.

The Gold Report for COPD outlines a method for distinguishing between of asthma, COPD, and asthma-COPD overlap syndrome (ACOS). While some of the indicators rely on PFT results, the Gold Report provides methods for making diagnoses without access to PFT [7]. However, this method relies on a trained pulmonologist who can recognize lung sounds and has familiarity with pulmonary diseases.

COPD and asthma are particularly hard to distinguish, because they share a variety of symptoms. This has led some researchers to believe that these diseases may both be part of a common etiology, the so-called “Dutch Hypothesis.” [8] This similarity presents challenges to both physicians and automated systems that are trying to diagnose or screen for these diseases.

C. The Opportunity for Automated and Mobile Diagnosis

The proliferation and availability of smart phones worldwide provides an affordable means to deploy diagnostic guidance tools at scale. While the cost of smart phone may yet be too high for individuals living in the poorest regions, the penetration of smart phones among rural doctors and clinical staff is quite high. These smart phones can be used as part of toolkits for screening and diagnosis.

When a trained pulmonologist is not available, machine learning algorithms running on the phone can be used by community health workers to screen for pulmonary disease, and refer specific patients for further testing at a specialized clinic. Such tools can also be used by general practitioner doctors to provide guidance and improve their diagnoses. Algorithms implemented on the mobile phone can combine information about symptoms, risk factors, abnormal lung sounds, and reduced lung function to identify possible diagnoses. Some prior work has been done in this area [9], [10], but the computational and hardware requirements have limited adoption in the areas where it can provide the most benefit.

Unlike a simple questionnaire, phones enable direct measurement of lung physiology, such as lung sounds or peak flow. All of these measurements can be analyzed at the point of care by the phone and diagnostic guidance can be provided to a physician or health worker. If connectivity is available, data can be transmitted to a remote pulmonologist who can review recommended diagnoses. If connectivity is not available, data can be stored and later reviewed by a pulmonologist to confirm diagnoses and identify patients who require follow-ups.

In this paper, we present a working design for a mobile platform for screening and diagnostic guidance for the top two chronic pulmonary diseases, COPD and asthma.

II. IMPLEMENTATION

A. Mobile Diagnostic Guidance Platform

The diagnostic algorithm presented in this paper makes use of a low-cost, mobile phone data collection tool published previously [11]. The platform is based on an Android



Figure 2: Peak flow meter and AR application

smartphone and consists of three components: a custom-built electronic stethoscope, a method of digitizing peak flow meter readings, and a patient questionnaire, described below.

1) Electronic Stethoscope

The construction and characterization of the electronic stethoscope was described previously [12]. In brief, the electronic stethoscope is created by modifying a traditional Microtone stethoscope. To add a recording capability, the eartips are removed and a microphone (CUI Inc. CMC-2742WBL-25L) is inserted into the tubing. The microphone is wired to a 3.5 mm Mono connection. The connection is connected to a splitter, which allows the user to record lung sounds and listen to them playback through headphones simultaneously. The stethoscope is shown in Figure 1.

2) Peak Flow Meter

Using Augmented Reality (AR) techniques implemented using the mobile camera, we developed an application that allows a user to automatically digitize their peak flow meter readings [13]. After the user aims their smartphone camera at the peak flow meter, the reading is automatically detected and recorded. The application enables longitudinal patient tracking and provides real-time feedback comparing the patient's readings to the population normal. The peak flow meter application is shown in Figure 2.

3) Questionnaire

To design the questionnaire, we reviewed previous efforts to produce symptom-based diagnostic guidelines [6], [7], [14]–[17]. These sources were used to create a questionnaire that captures information about the onset, duration, and progress of breathlessness, coughing, chest pain, fever, and nasal symptoms. The questionnaire also captures demographic information, medical history, and risk factor information such as smoking history or family history of allergies.

B. Diagnostic Algorithms

TABLE I. NUMBER OF PATIENTS WITH EACH DIAGNOSIS

Diagnosis	Number of Patients
Healthy	18
Asthma	48
COPD	17
Other Pulmonary Disease	16

Using data collected with the mobile diagnostic guidance platform, we trained a machine learning classification algorithm to identify patients with asthma and COPD. The algorithm was designed as a two-staged classifier. The first stage focused on separating individuals with asthma or COPD from the rest of the patient population. The second stage of the classifier focused on separating patients with asthma from patients with COPD. A two-stage approach was selected because it allows decision makers to more easily specify the acceptable trade-offs between false positives and false negatives. This would allow the user to require a higher burden of proof before declaring patients healthy vs. declaring a patient has asthma instead of COPD.

Both models were implemented as logistic regression models with L1 regularization. To select a value for the regularization parameter, we used leave-one-out cross validation. This method was chosen because it allowed us to identify only those features most important for accurate classification. This enables us to minimize the number of questions or tests required to identify patients with asthma or COPD.

The models were evaluated using leave-one-out cross validation. ROC curves were created by combing the class probabilities across the validation folds. The models were created using the SciKit-Learn Python Library 0.17.0 [18].

III. CLINICAL STUDY AND RESULTS

A. Data Collection

The data used for the present study is part of an ongoing study to develop screening algorithms for a variety of pulmonary diseases in rural India. The screening for COPD and asthma is the first phase of this study. The analysis described here uses data from the first 119 patients of the larger study. For each patient, a technician recorded lung sounds from eleven different recording locations. An experienced pulmonologist reviewed the recordings and indicated the presence of wheezes or crackles at each location. Each patient took three peak flow meter readings and all values were recorded. The technician administered the patient questionnaire and recorded the responses. 20 patients were excluded from the analysis because peak flow meter readings were not recorded.

To determine patient diagnoses, each patient underwent a typical clinical exam, followed by complete examination in the Pulmonary Function Testing lab (spirometry, body plethysmography, pulmonary gas diffusion, and impulse

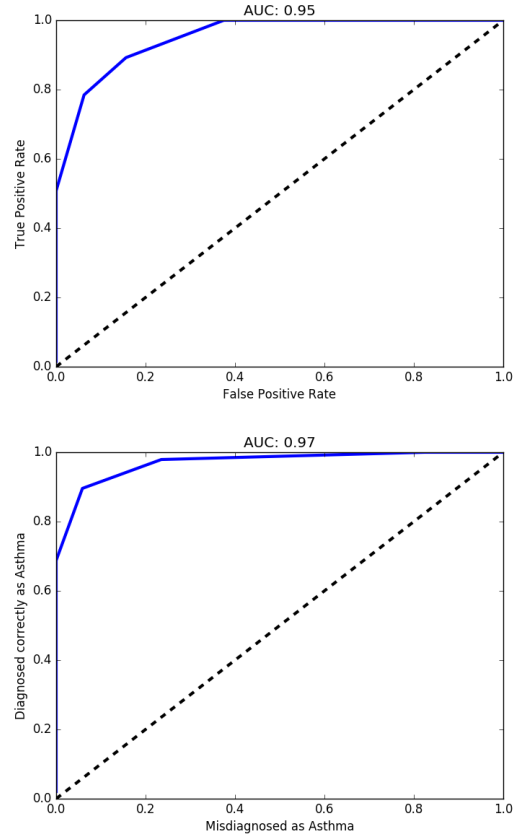


Figure 3: Receiver Operating Characteristic curves for the two classifiers: (top) Separating asthma/COPD from the general population (bottom) Separating asthma from COPD.

oscillometry). In order to identify allergic rhinitis, the pulmonologist followed ARIA guidelines [15]. Asthma and COPD were diagnosed according to ERS/ATS guidelines. The number of patients with each diagnosis is shown in Table I.

B. Algorithm Results

The first stage classifier had an AUC of 0.95. The ROC curve is shown in Figure 3. The features selected with L1 regularization were the presence of abnormal lung sounds, the presence of breathlessness or nasal symptoms, peak flow meter readings, and family or personal history with allergies.

The second stage classifier had an AUC of 0.97. The ROC curve is shown in Figure 3. The features selected with L1 regularization were the presence of abnormal lung sounds, peak flow meter readings, family history with allergies and patient's smoking history.

IV. DISCUSSION

Our classification algorithms were able to accurately identify patients with asthma and COPD from the general patient population. Each classifier needed only a few features to achieve a high level of accuracy.

One of the features in our diagnostic algorithm was whether or not the patient had abnormal lung sounds. Since a primary care physician or health worker may not be able to

TABLE II. STAGE I FEATURE COEFFICIENTS. POSITIVE VALUES INDICATE FEATURES ARE INDICATIVE OF ASTHMA AND COPD

Feature	Coefficient
Breathlessness	0.99
Family History of Allergies	0.17
Abnormal Lung Sounds	0.14
Nasal Symptoms	0.06
Personal History of Allergies	0.05
Maximum PFM Reading	-0.37
Maximum PFM Reading Compared to Reference	-1.01

accurately provide this data to the algorithm, we have separately developed algorithms to automatically identify abnormal lung sounds [19]. For this paper, the pulmonologist's identification of the sounds was used.

Many of the questions on our questionnaire were not found to be useful for distinguishing between asthma and COPD. If this result remains consistent as our study continues as we expand to additional diseases, it would justify the removal of these questions from the questionnaire.

One limitation of this analysis is that the accuracy of the models was evaluated using cross-validation. As data is collected from additional patients, we will define a held-out test set and use it to evaluate model performance.

V. CONCLUSION

We have demonstrated preliminary success in creating a mobile data collection platform and diagnostic algorithm that can successfully identify patients with COPD or Asthma. This work represents an early demonstration of an automated point-of-care platform for diagnosing pulmonary diseases using a mobile phone.

All physiological data used in the study was collected using our mobile phone tools, with a total cost less than US\$30, not counting the cost of the phone itself. While the current implementation of the algorithm was developed in Python and implemented on a PC computer, future versions of the mobile platform will be entirely implemented on the mobile phone and include screening for additional diseases.

Mobile diagnostic tools such as these serve an important role as screening tools for community health workers and diagnostic guidance for primary care doctors in low-resource areas.

REFERENCES

- [1] "WHO | The top 10 causes of death," *WHO*. [Online]. Available: <http://www.who.int/mediacentre/factsheets/fs310/en/>.
- [2] W. H. Organization, *Global Surveillance, Prevention and Control of Chronic Respiratory Diseases: A Comprehensive Approach*. World Health Organization, 2007.
- [3] R. Pellegrino, G. Viegi, V. Brusasco, R. O. Crapo, F. Burgos, R. Casaburi, A. Coates, C. P. M. van der 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," *Eur. Respir. J.*, vol. 26, no. 5, pp. 948–968, Nov. 2005.
- [4] J. Kaur, K. Chugh, A. Sachdeva, and L. Satyanarayana, "Under diagnosis of asthma in school children and its related factors," *Indian Pediatr.*, vol. 44, no. 6, pp. 425–428, Jun. 2007.
- [5] A. Mandke and K. Mandke, "Under diagnosis of COPD in primary care setting in Surat, India," *Eur. Respir. J.*, vol. 46, no. suppl 59, p. PA1809, Sep. 2015.
- [6] Knowledge Translation Unit: University of Cape Town LUNG INSTITUTE, "Primary Care 101," Health Department: Republic of South Africa, 2014.
- [7] R. A. Pauwels, A. S. Buist, P. M. A. Calverley, C. R. Jenkins, and S. S. Hurd, "Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Pulmonary Disease," *Am. J. Respir. Crit. Care Med.*, vol. 163, no. 5, pp. 1256–1276, Apr. 2001.
- [8] E. R. Bleeker, "Similarities and differences in asthma and COPD. The Dutch hypothesis," *Chest*, vol. 126, no. 2 Suppl, p. 93S–95S; discussion 159S–161S, Aug. 2004.
- [9] N. Gavriely, M. Nissan, D. W. Cugell, and A. H. Rubin, "Respiratory health screening using pulmonary function tests and lung sound analysis," *Eur. Respir. J.*, vol. 7, no. 1, pp. 35–42, Jan. 1994.
- [10] M. Yeginer, K. Ciftci, U. Cini, I. Sen, G. Kilinc, and Y. P. Kahya, "Using lung sounds in classification of pulmonary diseases according to respiratory subphases," in *26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2004. IEMBS '04*, 2004, vol. 1, pp. 482–485.
- [11] D. Chamberlain, R. Kodgule, and R. Fletcher, "Towards a Pulmonary Diagnostic Kit for Telemedicine and Global Health Point-Of-Care Diagnosis," in *NIH-IEEE 2015 Strategic Conference on Healthcare Innovations and Point-of-Care Technologies for Precision Medicine*, 2015.
- [12] D. Chamberlain, R. Kodgule, J. Mofor, and R. Fletcher, "Mobile Stethoscope and Signal Processing Algorithms for Pulmonary Screening and Diagnostics," in *2015 IEEE Global Humanitarian Technology Conference (GHTC)*, 2015.
- [13] D. Chamberlain, R. Kodgule, A. Jiménez-Galindo, and R. Fletcher, "Applying Augmented Reality on Mobile Phones to Enable Automated and Low-Cost Data Capture from Medical Devices," in *Proceedings of the Eighth International Conference on Information and Communication Technologies and Development*, Ann Arbor, MI, 2016.
- [14] R. G. English, M. O. Bachmann, E. D. Bateman, M. F. Zwarenstein, L. R. Fairall, A. Bheekie, B. P. Majara, C. Lombard, R. Scherpbier, and E. S. Ottomani, "Diagnostic accuracy of an integrated respiratory guideline in identifying patients with respiratory symptoms requiring screening for pulmonary tuberculosis: a cross-sectional study," *BMC Pulm. Med.*, vol. 6, p. 22, 2006.
- [15] "Management of Allergic Rhinitis and Its Impact on Asthma: Pocket Guide," Allergic Rhinitis and Its Impact on Asthma, 2008.
- [16] "Respiratory Illness in Children and Adults, Diagnosis and Treatment of," Institute for Clinical Systems Improvement, 2013.
- [17] "Chronic obstructive pulmonary disease in over 16s: diagnosis and management | Guidance and guidelines | NICE." [Online]. Available: <http://www.nice.org.uk/guidance/cg101>. [Accessed: 21-Feb-2016].
- [18] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay, "Scikit-learn: Machine Learning in Python," *J Mach Learn Res*, vol. 12, pp. 2825–2830, Nov. 2011.
- [19] D. Chamberlain, R. Kodgule, D. Ganelin, V. Miglani, and R. R. Fletcher, "Application of Semi-Supervised Deep Learning to Lung Sound Analysis," in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Submitted.