Predicting Spirometry Readings Using Cough Sound Features and Regression

Roneel V. Sharan*, Udantha R. Abeyratne, Vinayak R. Swarnkar, Scott Claxton, Craig Hukins, and Paul Porter

Abstract

Objective: Spirometry is a commonly used method of measuring lung function. It is useful in the definitive diagnosis of diseases such as asthma and chronic obstructive pulmonary disease (COPD). However, spirometry requires cooperative patients, experienced staff, and repeated testing to ensure the consistency of measurements. There is discomfort associated with spirometry and some patients are not able to complete the test. In this paper, we investigate the possibility of using cough sound analysis for the prediction of spirometry measurements.

Approach: Our approach is based on the premise that the mechanism of cough generation and the forced expiratory maneuver of spirometry share sufficient similarities enabling this prediction. Using an iPhone, we collected mostly voluntary cough sounds from 322 adults presenting to a respiratory function laboratory for pulmonary function testing. Subjects had the following diagnoses: obstructive, restrictive, or mixed pattern diseases, or were found to have no lung disease along with normal spirometry. The cough sounds were automatically segmented using the algorithm described in [1]. We then represented cough sounds with various cough sound descriptors and built linear and nonlinear regression models connecting them to spirometry parameters. Augmentation of cough features with subject demographic data is also experimented with. The dataset was divided into 272 training subjects and 50 test subjects for experimentation.

Main Results: The performance of the auto-segmentation algorithm was evaluated on 49 randomly selected subjects from the overall dataset with a sensitivity and PPV of 84.95% and 98.51%, respectively. Our regression models achieved a root mean square error (and correlation coefficient) for standard spirometry parameters FEV1, FVC, and FEV1/FVC of 0.593L (0.810), 0.725L (0.749), and 0.164 (0.547), respectively, on the test dataset. In addition, we could achieve sensitivity, specificity, and accuracy of 70% or higher by applying the GOLD standard for COPD diagnosis on the estimated spirometry test results.

Significance: The experimental results show high positive correlation in predicting FEV1 and FVC and moderate positive correlation in predicting FEV1/FVC. The results show possibility of predicting spirometry results using cough sound analysis.

Keywords

Cough sound analysis, chronic respiratory disease, regression, spirometry.

I. Introduction

Globally, preventable chronic respiratory diseases affect hundreds of millions of people. In 2015, chronic obstructive pulmonary disease (COPD) was among the top four causes of death

This work was supported in part by ResApp Health Limited and The University of Queensland.

*R. V. Sharan, U. R. Abeyratne and V. R. Swarnkar are with the School of Information Technology and Electrical Engineering, The University of Queensland, Brisbane, QLD 4072, Australia (e-mail: r.sharan@uq.edu.au).

S. Claxton is with GenesisCare and Joondalup Health Campus in Perth, Australia.

C. Hukins is with Princess Alexandra Hospital in Brisbane, Australia.

P. Porter is with the Princess Margaret Hospital, Joondalup Health Campus, and Faculty of Health Sciences, Curtin University, in Perth, Australia.

worldwide (COPD and asthma accounted for approximately 6.3% of the deaths [2]). According to the Australian Bureau of Statistics, chronic lower respiratory diseases [3], which includes diseases such as COPD and asthma, were the fifth largest cause of death in Australia in 2015, accounting for about 5% of all deaths [4].

Early diagnosis of chronic respiratory diseases is important to reduce its severity and effects. The stepwise investigation into the diagnosis of chronic respiratory diseases proposed by the Global Alliance against Chronic Respiratory Diseases (GARD) includes lung function measurement [5]. Lung function tests (LFTs), also known as pulmonary function tests (PFTs), allow for a quantifiable assessment of pulmonary function without physical examination of the lungs. Spirometry, performed using a spirometer, is the most common form of LFT [6]. It is a powerful tool in the diagnosis and management of respiratory diseases [7] and referred as the gold standard in the diagnosis and assessment of COPD [5].

Spirometry measures the volume and/or speed of air that can be inhaled and exhaled. During spirometry, a subject places their mouth around a mouthpiece of a spirometer and takes the deepest possible breath. A forced expiration is then performed with the target of expelling all the air as rapidly as physically possible [8]. It is essential that the patient fully engages with the process and exerts the maximum effort possible. The procedure is repeated until three consistent measurements are obtained. The spirometer measures the forced expiratory volume in the first second (FEV1) and the forced vital capacity (FVC) which represents the total volume of air exhaled from the lungs after the deepest possible inhalation.

The spirometry results are often compared against the predicted or reference values [9]. The forced expiratory ratio, FEV1/FVC, and FVC help differentiate obstructive, restrictive, and normal breathing patterns [7]. Abnormal results could indicate obstructive respiratory conditions, such as asthma and COPD, and restrictive diseases, such as interstitial lung disease and obesity [7, 10, 11]. In addition, the severity of the obstructive disease can be determined using FEV1.

Although spirometry is a non-invasive test, it requires significant patient cooperation and also physical contact. The latter requires contact via a mouthpiece into the spirometer and often also a nose clip in order to minimize air loss. The test requires subjects to understand and cooperate in exhaling as hard and fast as possible. The test may need to be repeated multiple times to ensure consistency in results [7, 8]. This can be inconvenient and difficult for some subjects such as the elderly or those with significant lung disease. Additionally, the cost of equipment and personnel time is significant. Also, more than 50% of people with chronic respiratory diseases live in low and middle income countries [5] where medical resources and expertise is scarce. This brings about the need for new techniques and readily available devices for performing such tests without the need for significant medical expertise.

In comparison, the recording of cough sounds requires minimal patient cooperation (particularly spontaneous coughs; and voluntary coughs can usually be easily produced by adults) and requires no physical contact. Cough is a common symptom of various respiratory conditions. It is a natural reflex of the human body to clear the throat or airways of foreign particles or mucus.

The phases of spirometry and cough physiology share some similarities. Cough physiology is comprised of three phases: inspiratory, compression, and expiratory [12]. During inspiration, an amount of air is inhaled which can be seen as similar to the first step of spirometry where the subject takes a deep breath. Compression is a brief period when the glottis closes, maintaining lung volume as intrathoracic pressure builds. In spirometry, this phase could be seen as similar to

the subject holding their breath before expiration. The expiratory phase of a cough starts with the rapid opening of the glottis. There is an initial supramaximal expiratory flow of air followed by a longer lasting lower expiratory flow [12-14]. This compares to the last step of spirometry where the subject expels the bulk of the air initially followed by reduced but prolonged expulsion.

These similarities lead to the possibility that spirometry results can be estimated using cough sound analysis. In previous studies, cough sound signal classification has been shown to be useful in detecting respiratory diseases such as pneumonia [15], asthma [16], and croup [1]. In a recent work [17], cough sounds have been used to estimate spirometry readings for asthmatic and healthy subjects. The study reported root mean square error (RMSE) of 0.48L, 0.57L, and 0.08 in estimating FEV1, FVC, and FEV1/FVC, respectively. The study dataset, however, was small with only 28 subjects: 16 healthy and 12 asthmatic. In addition, other obstructive, restrictive, and mixed diseases were not explored and the cough signals were manually segmented.

In this study, we explore the use of cough sound signal analysis for estimating spirometry readings. Compared to [17], our technique utilizes various cough sound descriptors. In addition, our dataset is significantly larger, at a total of 322 subjects, and our methods are fully automated starting from cough episode extraction. Along with asthma and subjects with normal LFT, we also include subjects with COPD, restrictive, and mixed lung diseases. In addition, we employ sequential backward feature selection (SBFS) in an effort to improve the prediction. Our paper explores linear and nonlinear regression models.

Moreover, age, height, gender, and ethnicity of the subject are the main determinants in the predicted/reference spirometry values [10, 18]. The weight of the subject may also be used in some reference equations [8]. The augmentation of cough features with subject demographic data has been shown to be useful in classifying respiratory diseases [19, 20]. In this work, we test the usefulness of subject demographic information in predicting the spirometry readings when augmented with the cough sound features.

The rest of this paper is organized as follows. The data collection method is described in section II. Technical details of the proposed approach are presented in section III. The dataset and model development and testing are discussed in section IV and conclusions in section V.

II. Data Collection

The human ethics committees of The University of Queensland, Brisbane, Australia, and Joondalup Health Campus (JHC), Perth, Australia, approved the study protocols and patient recruitment procedure. The study population consists of subjects presenting with respiratory symptoms including cough, sputum production, wheeze, and shortness of breath and who were undergoing spirometry testing in a dedicated pulmonary function facility at JHC.

The cough sounds were recorded within 15 minutes of performing the spirometry test. The recording was done using an Apple iPhone 6s and with help of an iOS application developed by ResApp Health limited. Sound data was recorded at a sampling rate of 44100 samples per second at a bit depth of 16 bits per sample. The smartphone recorder was placed approximately 20-50 cm away from the mouth of the subject at an angle of approximately 45°. This helped in eliminating the effects of wind noise which arises when bulk air streams of a cough directly hit the microphone.

Cough recordings were made by dedicated research nurses in the realistic environment of a

pulmonary function laboratory. Our protocols attempted to eliminate preventable interferences such as loud conversations, overlapping coughs, and coughs from parties other than the targeted subject. However, some of the recordings contained unavoidable interferences such as speech sounds and beeps and noises contributed by medical devices being used in the environment at the time of cough recording.

III. Proposed Method

The main steps in the proposed approach are shown in Fig. 1 [17]. The dataset is divided into training and test subjects. The regression model is trained and validated with features extracted from all coughs from all training subjects. The trained/validated model is then used to estimate the spirometry readings for all coughs for the test subject. The final estimation for each test subject is statistically determined from the predicted cough outputs.

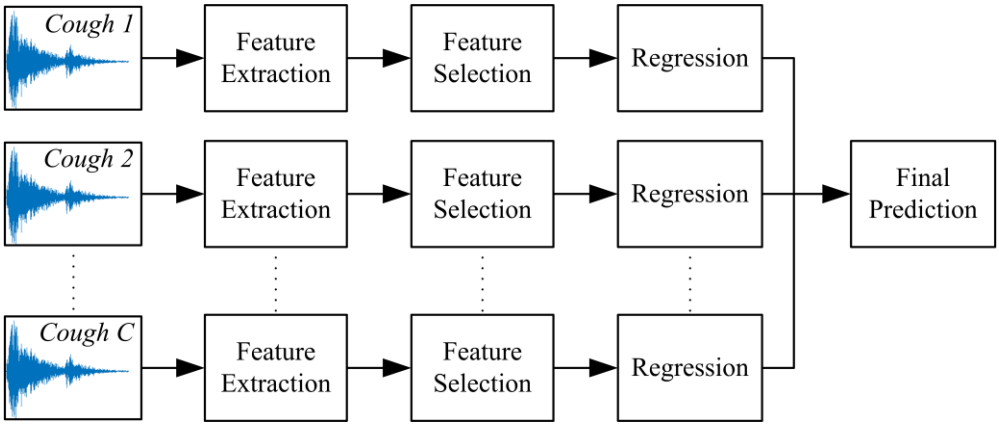

Fig. 1. Procedure for testing the trained regression model.

A. *Features*

The cough sound signals are automatically segmented using the method described in [1, 21]. The segmented coughs are then divided into three equal parts and various features are extracted from each segment. These are $8\times$ bispectrum scores, $1\times$ non-Gaussianity score, $4\times$ formants (first four formants), $1\times$ log energy, $1\times$ Shannon's entropy, $1\times$ zero-crossing rate, $1\times$ kurtosis, and $31\times$ mel-frequency cepstral coefficients (MFCC). In addition, $13\times$ wavelet features are extracted from the cough signal. This gives a final cough feature dimension of 157. More details on the usefulness of these features in cough sound analysis and feature extraction can be found in [15, 19].

In the cough + demographic feature models, subject demographic data, age, gender, weight, and height, are also utilized.

B. *Regression*

In this work, we experiment with one linear and one non-linear regression method, linear regression [22] and support vector regression (SVR) [23], respectively. We use the least-squares fit for linear regression and nonlinear regression with Gaussian kernel for SVR.

Regression is used to predict the output of each cough sound signal. We experiment with *mean*,

median, *minimum*, and *maximum* statistics to predict the final output for each test subject from the corresponding cough predicted values (see Fig. 1).

C. Feature Selection

The aim of sequential backward feature selection is to remove irrelevant feature dimensions and select a subset of features that minimize the RMSE. The following steps are followed in SBFS.

Step 1: Use all the feature dimensions to calculate the baseline RMSE.

Step 2: One feature dimension is removed at a time and the RMSE is calculated with the remaining features. At the end of this step, the feature removal corresponding to the lowest RMSE is permanently removed and the feature set is updated.

Step 3: Repeat step 2 with the remaining features. This process continues until no further improvement can be achieved in the prediction performance.

Step 4: The SBFS algorithm terminates when the prediction error plateaus and the remaining features are utilized in training and testing the final models.

IV. Model Training and Evaluation

In this section, we first give an overview of the dataset used in this study. We then discuss our algorithm development and validation procedures. This is followed by the results obtained during model training, validation, and testing.

A. Database Overview

Our database consists of cough sound recordings and detailed clinical diagnostic information on each subject including the final diagnosis, clinical examination findings, LFT outcomes, and subject demographic information.

The cough sound database has a total of 322 adult subjects with obstructive, restrictive, and mixed (obstructive and restrictive) respiratory conditions and subjects with normal LFT results. It also includes some subjects with non-chronic respiratory diseases. These subjects were initially suspected of having chronic respiratory conditions and, therefore, were subjected to the spirometry procedure. The obstructive group has been divided into three subgroups for further analysis. These are COPD, asthma, and other obstructive diseases (diseases which could not be grouped into COPD and asthma such as COPD and asthma comorbids, bronchiectasis, emphysema, etc.). Each diagnosis was made by a specialist respiratory physician and confirmed using formal lung function tests.

The dataset is divided into training and test sets of 272 subjects and 50 subjects, respectively. The regression model is trained and validated on the training dataset. The test dataset is completely independent of the training dataset and is used for testing only. The cough dataset has multiple cough sounds (mostly voluntary coughs) recorded for each subject. The cough sound signals are automatically segmented the procedure for which is described in [1, 21]. Breakdown of demographic and cough data for the training (and test) subjects are given in Table I.

Table I: Statistical overview of demographic and cough data for training (and test) subjects

	Number of Subjects	Average Age (Years)	Gender (M:F)	Height (m)	Weight (kg)	Cough Data		Smoking History (Yes:No)
						Total No. of Coughs	Average No. of Coughs/ Subject	
Obstructive (COPD)	52 (10)	72 ± 9 (69 ± 10)	22:30 (3:7)	1.65 ± 0.09 (1.65 ± 0.13)	73.72 ± 15.92 (67.32 ± 19.47)	479 (95)	9.21 (9.50)	49:3 (9:1)
Obstructive (Asthma)	39 (8)	66 ± 16 (68 ± 12)	20:19 (5:3)	1.69 ± 0.11 (1.69 ± 0.09)	83.31 ± 18.61 (84.89 ± 10.20)	383 (80)	9.82 (10.00)	20:19 (4:4)
Obstructive (Others)	19 (5)	68 ± 16 (62 ± 6)	7:12 (3:2)	1.65 ± 0.08 (1.68 ± 0.09)	74.27 ± 15.29 (87.00 ± 21.05)	189 (50)	9.95 (10.00)	10:9 (4:1)
Restrictive	42 (11)	68 ± 15 (63 ± 20)	26:16 (10:1)	1.67 ± 0.10 (1.74 ± 0.07)	85.33 ± 19.58 (92.94 ± 16.46)	412 (94)	9.81 (8.55)	25:17 (8:3)
Mixed Pattern	6 (0)	62 ± 17 (-)	4:2 (-)	1.63 ± 0.14 (-)	101.58 ± 24.22 (-)	65 (-)	10.83 (-)	4:2 (-)
LFT Normal	83 (16)	63 ± 15 (59 ± 14)	31:52 (8:8)	1.67 ± 0.10 (1.69 ± 0.08)	80.29 ± 17.83 (83.91 ± 17.82)	824 (128)	9.93 (8.00)	43:40 (7:9)
Others	31 (0)	67 ± 11 (-)	16:15 (-)	1.68 ± 0.11 (-)	80.85 ± 14.44 (-)	313 (-)	10.10 (-)	21:10 (-)
Overall	**272 (50)**	**67 ± 14 (64 ± 14)**	**126:146 (29:21)**	**1.67 ± 0.10 (1.69 ± 0.09)**	**80.36 ± 18.06 (83.04 ± 18.65)**	**2665 (447)**	**9.80 (8.94)**	**172:100 (32:18)**

B. Experimental Setup

Algorithm development and evaluation was carried out as described in Section III.

Background noise and unavoidable interferences were part of the natural environment in which the recordings were made. We applied a 4^{th} order Butterworth bandpass filter, lower cutoff frequency of 70Hz and upper cutoff frequency of 20kHz, to the cough sound signals to filter out low frequency noise. In addition, we use all available coughs to predict the final spirometry reading for a subject. This minimizes the effect that coughs corrupted with noise might have on the final estimated value.

We used the leave-one-subject-out validation technique to train and validate our models. That is, all cough sound signals from a single subject are used for testing and all cough sound signals from all other subjects are used for training the regression model, making the trained model independent of the test subject. This process is repeated for all subjects resulting in the number of trained models equal to the number of subjects. Leave-one-subject-out validation is performed on the training dataset only. The model is then retrained using all the training observations and with the optimal settings determined during validation for testing on the independent test dataset.

Using the laboratory spirometry measurements as reference, we evaluate the prediction performance of the models using the RMSE and the standard deviation of the RMSE, as defined in [17]. For SVR, all results are reported using nonlinear SVM with a Gaussian kernel. SVM parameters, the penalty parameter and the width of the Gaussian function [24], were tuned using Bayesian optimization [25]. The aim in parameter tuning was to minimize the RMSE of leave-one-subject-out validation.

C. Performance of the Automatic Segmentation Algorithm

The automatic segmentation algorithm and its training and validation procedures have been described in [1, 21] and is summarized here. The auto segmentation algorithm utilizes Time-Delay Deep Neural Network (TD-DNN) with autoencoders in the hidden layers [26, 27]. The 37-dimensional input feature vector for the auto segmentation algorithm includes cepstral, entropy, and temporal features. Using variance hypothesis test, a total of 25 features were determined to be useful in representing cough sounds for segmentation. The algorithm was trained and validated on a dataset which is totally independent of the dataset used in this work. The training/validation dataset had a total of 153 subjects with a sensitivity and positive predictive value (PPV) of 89.79% and 80.55% in leave-one-out cross-validation [1].

We tested the performance of the auto segmentation algorithm on a subset of the dataset used in this work. Cough sound signals were manually segmented so that the start and end points can be compared as per the procedure described in [21]. Manual segmentation is a time consuming process, therefore, coughs from only 49 randomly selected subjects was used for this purpose. The cough sounds for these subjects were then segmented using the automatic segmentation algorithm. A sensitivity and PPV of 84.95% and 98.51% was achieved with the manually segmented coughs as reference.

D. Results Using Linear Regression

1) Results Using Cough Feature Model

The RMSE in predicting FEV1, FVC, and FEV1/FVC using cough only features are given in Table II. The results are before and after feature selection and the mean, median, minimum, and maximum statistical measures are used in predicting the final spirometry readings.

The RMSE for FEV1, FVC, and FEV1/FVC using mean measure are 0.739L, 0.895L, and 0.150, respectively. This improves to 0.688L, 0.832L, and 0.138 after feature selection. Feature selection is seen to improve the prediction error for all models. With the mean measure, the improvement in the RMSE after feature selection is 0.051L, 0.063L, and 0.012 for FEV1, FVC, and FEV1/FVC, respectively. Feature selection is also seen to reduce the standard deviation of the RMSE by 0.075L, 0.090L, and 0.004 for FEV1, FVC, and FEV1/FVC, respectively.

The mean and median statistical measures are seen to give the best prediction performance for the cough only model. In general, only a marginal difference is observed between the prediction errors for the median and mean measures. Unlike the median measure which picks one value from set of predicted outcomes, the mean has the advantage of using all the values in the computation of the final prediction. Therefore, from now on we use the mean measure for final spirometry reading prediction.

Table II: RMSE ± standard deviation of the RMSE for predicting FEV1, FVC, and FEV1/FVC on the training/validation dataset. The model utilizes cough features and linear regression.

	Before Feature Selection			After Feature Selection		
Final Prediction Method	**FEV1 (L)**	**FVC (L)**	**FEV1/FVC**	**FEV1 (L)**	**FVC (L)**	**FEV1/FVC**
Mean	0.739 ± 0.814	0.895 ± 1.092	0.150 ± 0.031	0.688 ± 0.739	0.832 ± 1.002	0.138 ± 0.027
Median	0.744 ± 0.830	0.890 ± 1.102	0.150 ± 0.032	0.684 ± 0.734	0.825 ± 0.974	0.137 ± 0.026
Min	0.979 ± 2.053	1.132 ± 1.896	0.179 ± 0.064	0.810 ± 1.133	0.986 ± 1.553	0.152 ± 0.029
Max	1.186 ± 5.949	1.415 ± 9.438	0.179 ± 0.051	0.829 ± 0.745	0.996 ± 1.202	0.152 ± 0.035

2) Results Using Cough + Demographic Feature Model

The RMSE using cough + demographic feature model are presented in Table III. The demographic data includes age, gender, weight, and height of the subject. The results are once again presented before and after feature selection. The RMSE for the cough + demographic feature model shows significant improvement when compared to the corresponding results for the cough feature model. For the before feature selection models, the improvement in the RMSE for FEV1, FVC and FEV1/FVC are 0.173L, 0.220L, and 0.005, respectively. A similar observation is also made for the cough + demographic feature model and cough feature model after feature selection. The improvement in RMSE for FEV1, FVC, and FEV1/FVC are 0.164L, 0.210L, and 0.005, respectively.

There is also significant reduction in the standard deviation of the RMSE with the introduction of the demographic features. The reduction in the standard deviation for FEV1, FVC, and FEV1/FVC for the before feature selection models are 0.316L, 0.164L, and 0.001, respectively. For the two after feature selection models, the reduction in the standard deviation for FEV1, FVC, and FEV1/FVC are 0.306L, 0.225L, and 0.001, respectively. Therefore, the augmentation of demographic information to the cough sound descriptors has shown to be useful in estimating the spirometry readings.

In addition, feature selection has once again shown to be useful in reducing the RMSE. For the cough + demographic feature model, the RMSE (and standard deviation of the RMSE) is seen to improve by 0.042L (0.065L), 0.053L (0.151L), and 0.012 (0.004) for FEV1, FVC, and FEV1/FVC, respectively.

Table III: RMSE ± standard deviation of the RMSE for predicting FEV1, FVC, and FEV1/FVC on the training/validation dataset. The model utilizes cough + demographic features and linear regression.

Feature Set	**FEV1 (L)**	**FVC (L)**	**FEV1/FVC**
Before Feature Selection	0.566 ± 0.498	0.675 ± 0.928	0.145 ± 0.030
After Feature Selection	0.524 ± 0.433	0.622 ± 0.777	0.133 ± 0.026

3) Test Results

The test results using linear regression are given in Table IV. Only the cough + demographic feature models were used here since it has shown to be more accurate in estimating the spirometry readings on the training/validation dataset. Interestingly, the RMSE for FEV1 and FVC using the before feature selection model are better than the after feature selection model. In addition only a marginal improvement is seen in the RMSE for FEV1/FVC in the after feature selection model. This suggests that while feature selection could be used to improve the estimation performance on the validation dataset, the performance does not necessarily translate to the test dataset.

In general, it could be said that the before feature selection model gives the best estimation. The RMSE is estimating FEV1, FVC, and FEV1/FVC using this model are 0.630L, 0.750L, and 0.157, respectively.

Table IV: RMSE ± standard deviation of the RMSE for predicting FEV1, FVC, and FEV1/FVC on the test dataset. The model utilizes cough + demographic features and linear regression.

Feature Set	**FEV1 (L)**	**FVC (L)**	**FEV1/FVC**
Before Feature Selection Model	**0.630 ± 0.535**	**0.750 ± 0.678**	0.157 ± 0.034
After Feature Selection Model	0.631 ± 0.503	0.770 ± 0.715	**0.155 ± 0.033**

E. Results Using Support Vector Regression

We now present results using support vector regression. Since the test results using linear regression don't show any significant benefit of feature selection, in this instance we present results without feature selection.

1) Validation Results

The RMSE in predicting FEV1, FVC, and FEV1/FVC using cough feature model and cough + demographic feature model are given in Table V. The results use the mean statistical measure for final prediction. The results using SVR generally show improvement over the corresponding validation results using linear regression. The improvement in the cough feature model is 0.020L, 0.025L, and 0.006 for FEV1, FVC, and FEV1/FVC, respectively. Similarly, the improvement in the cough + demographic feature model is 0.016L, 0.016L, and -0.005, for FEV1, FVC, and FEV1/FVC, respectively.

Similar to linear regression, the RMSE for SVR shows improvement with the inclusion of demographic data. The improvement in the RMSE from the cough feature model to the cough + demographic feature model is 0.169L, 0.211L, and 0.006, respectively.

Table V: RMSE ± standard deviation of the RMSE for predicting FEV1, FVC, and FEV1/FVC on the training/validation dataset. The model utilizes cough features and cough + demographic features with support vector regression but without feature selection.

Model	**FEV1 (L)**	**FVC (L)**	**FEV1/FVC**
Cough Feature Model	0.719 ± 0.746	0.870 ± 1.080	0.144 ± 0.029
Cough + Demographic Feature Model	0.550 ± 0.552	0.659 ± 1.011	0.138 ± 0.030

2) Test Results

The test results using the cough + demographic feature model for SVR are given in Table VI. In this instance, we present the overall RMSE and the RMSE for the individual disease groups. The overall RMSE for FEV1 and FVC show improvement over the corresponding results using linear regression. The RMSE for FEV1/FVC is, however, slightly degraded. As far as the RMSE for the individual disease groups is concerned, FEV1 and FEV1/FVC are best estimated for restrictive diseases and FVC best estimated for asthma.

Table VI: RMSE ± standard deviation of the RMSE for predicting FEV1, FVC, and FEV1/FVC on the test dataset. The model utilizes cough + demographic features and support vector regression without feature selection.

Disease	**FEV1 (L)**	**FVC (L)**	**FEV1/FVC**
Obstructive (COPD)	0.904 ± 0.985	0.872 ± 0.781	0.280 ± 0.053
Obstructive (Asthma)	0.452 ± 0.189	**0.573 ± 0.497**	0.211 ± 0.037
Obstructive (Others)	0.522 ± 0.241	0.696 ± 0.453	0.082 ± 0.005
Restrictive	**0.442 ± 0.184**	0.722 ± 0.344	**0.061 ± 0.004**
LFT Normal	0.518 ± 0.241	0.701 ± 0.834	0.090 ± 0.008
Overall	0.593 ± 0.535	0.725 ± 0.670	0.164 ± 0.041

The regression plots for the corresponding test results are shown in Fig. 2. Pearson's correlation coefficient, *R*, of 0.810, 0.749, and 0.547 is achieved in predicting FEV1, FVC, and FEV1/FVC, respectively. As per [28], this shows a high positive correlation in predicting FEV1 and FVC and moderate positive correlation in predicting FEV1/FVC using cough sound descriptors and demographic data.

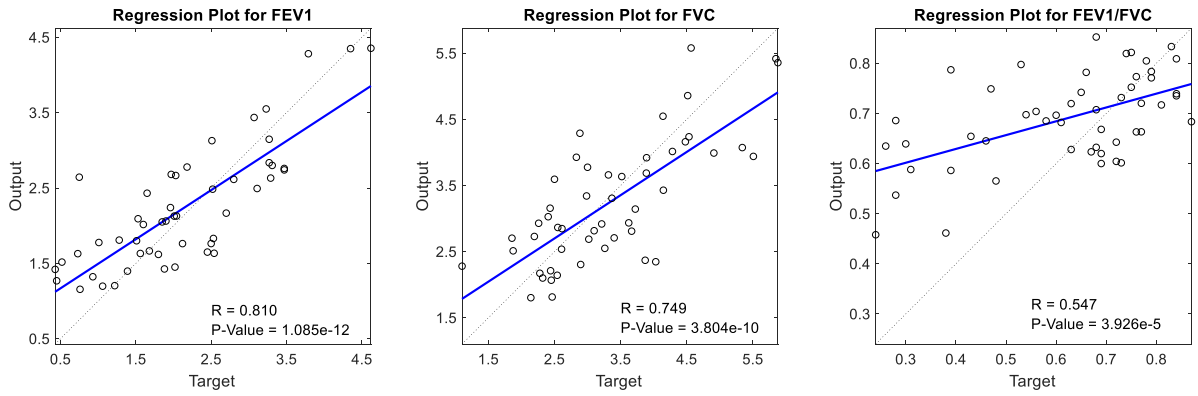

Fig. 2. Regression plot for the cough + demographic feature model as tested on the test dataset using support vector regression.

According on the GOLD standard [29], an important criterion in the diagnosis of COPD is that FEV/FVC < 0.7. We apply this criterion on the test dataset, the corresponding classification results for which are given in Table VII. We first apply this criterion to the laboratory spirometry measurements which serves as our reference diagnosis. We then apply the same to the spirometry results estimated by the regression models. The cough + demographic feature models for both linear and support vector regression are used for this purpose, the results for which are given in Table IV and Table VI, respectively. Using both regression models, we could achieve sensitivity, specificity, and accuracy of 70% or higher.

Table VII: Classification results using the GOLD criterion (FEV1/FVC < 0.7) on the test dataset. The model utilizes cough + demographic features and both linear and support vector regression.

Model	Sensitivity (%)	Specificity (%)	Accuracy (%)
Linear Regression	73.33	70.00	72.00
Support Vector Regression	70.00	70.00	70.00

V. Limitations

During cough sound recording, having the recorder directly in front of the mouth and too close to the mouth resulted in wind noise due to the rapid expulsion of air during cough. Placing the recorder at approximately 45° and 20-50cm away from the mouth of the subject minimized this problem. This was part of the recording guidelines provided to the clinicians performing the recording. We expect the models to behave as per the test results as long as the approximate recording device placement guidelines (angle and distance, in particular) are followed as slight variations of these have already been captured through the training data used in our models.

In addition, currently, the method is applicable to iPhone only. If using an iPhone and as per the guidelines then no extra calibration would be needed. However, we believe it will work equally well with other sound recording devices with similar characteristics. The smart phones of today have high fidelity and high bandwidth that were unthinkable a decade ago. The majority of features in our method are purpose-designed to be independent of the amplitude of cough

sounds. Therefore, slight variations of position or microphone characteristics are not expected to affect the results significantly.

We did not note the exact placement of the recording device during data collection and at this stage it is not known how the results may be affected if the recording guidelines are not strictly adhered to. Also, it is not known how recording devices with substantially different microphone characteristics will affect the performance of the models. These require further investigation which may bring about the need for feature calibration.

In the clinical database of 322 subjects used in this study, we had only 6 subjects presenting with obstructive and restrictive disease comorbidity (mixed pattern). The small number of subjects makes it difficult to train specific diagnostic models targeting the mixed pattern. Thus, in this paper we resort to reporting the performance of our algorithm of the mixed pattern subjects as part of the overall model validation procedure. The same process was followed in the case of the group "other" (refer to Table I) which comprises a number of disease groups, each of which carries insufficient numbers for a targeted model training. We hope to train targeted models on these groups when datasets become large enough.

The study population used in this study was subjects undergoing routine clinical spirometry testing in a pulmonary function laboratory of a hospital. The inclusion criteria was the existence of respiratory symptoms including cough, sputum production, wheeze, and shortness of breath. The validity of the results we present in this paper is currently limited to lung function laboratory cohorts. It is not known how coughs generated by conditions such as sleep apnea and post-nasal drip would affect the performance of the method. We note, however, while coughs due to benign reasons were not specifically part of the study, it included a number of subjects (99 subjects) whose spirometry results were normal.

VI. Discussion and Conclusion

Results on predicting spirometry readings using cough sound descriptors, demographic data, and regression are presented in this paper. All available coughs from each subject were used in estimating the spirometry readings. We also experimented with using a smaller number of coughs from each subject. However, using smaller number of coughs does not change the outcomes but was seen as a waste of opportunity since we had access to multiple coughs from each subject. The ability to use more coughs is likely to give an estimate with a lower variance. Using all available coughs to make the prediction gives greater confidence in the predicted value. Feature selection is a key component of classification and regression tasks and in this work we experimented with the sequential backward feature selection strategy. The validation error was seen to improve significantly with feature selection but, interestingly, the performance did not always translate to the independent test dataset. Apparently, feature selection bias is not a new problem and has been seen to be more evident in regression than classification [30].

The coefficient of correlation indicates high to moderate positive correlation between cough sounds and spirometry readings. Our results provide strong evidence to support the hypothesis that cough sounds carry sufficient information to estimate spirometry parameters. This work corroborates the initial findings reported in [17] on a small group of normal subjects and asthma patients. Our study uses a significantly larger data set and extends the study to represent a more realistic group of diseases mimicking the scenario found in a typical respiratory function laboratory. In addition, the method we developed is fully automated and can use an iPhone as a

sound recorder, computing device as well as the display interface. This greatly facilitates the translation of the technology outside of a pulmonary function laboratory and into ambulatory clinical settings.

An important question that arises is how would the diagnostic results using the estimated spirometry measurements compare to that based on the laboratory spirometry measurements. This is a difficult question to answer since the actual clinical diagnosis procedure in this study utilizes a combination of lung function test outcomes, clinical history of the patient and other clinical examination results as well as laboratory results, as appropriate. For this reason, we are unable to report on the actual diagnostic outcomes directly without running a separate outcome study. In order to overcome this problem, we reported the would be clinical diagnosis if the diagnosis were based solely on the GOLD standard (using spirometry measurements) and compared the performance that would result if we used our cough-based estimations in place of spirometry.

References

[1] R. V. Sharan, U. R. Abeyratne, V. R. Swarnkar, and P. Porter, "Automatic croup diagnosis using cough sound recognition," *IEEE Transactions on Biomedical Engineering*, 2018.

[2] "Global health estimates 2015: Deaths by cause, age, sex, by country and by region, 2000-2015," World Health Organization, Geneva, 2016.

[3] International statistical classification of diseases and related health problems (10th revision). *World Health Organization*. Available: www.who.int

[4] "Causes of Death, Australia, 2015," Australian Bureau of Statistics, 2015.

[5] "Global surveillance, prevention and control of chronic respiratory diseases: A comprehensive approach," World Health Organization, Switzerland, 2007.

[6] D. Brazzale, G. Hall, and M. P. Swanney, "Reference values for spirometry and their use in test interpretation: A position statement from the Australian and New Zealand society of respiratory science," *Respirology*, vol. 21, no. 7, pp. 1201-1209, Oct 2016.

[7] J. D. Johnson and W. M. Theurer, "A stepwise approach to the interpretation of pulmonary function tests," *American Family Physician*, vol. 89, no. 5, pp. 359-66, 2014.

[8] V. C. Moore, "Spirometry: step by step," *Breathe*, vol. 8, no. 3, pp. 232-240, 2012.

[9] P. H. Quanjer, S. Stanojevic, T. J. Cole, X. Baur, G. L. Hall, B. H. Culver, *et al.*, "Multi-ethnic reference values for spirometry for the 3-95-yr age range: The global lung function 2012 equations," *European Respiratory Journal*, vol. 40, no. 6, pp. 1324-1343, Dec 2012.

[10] D. Burton, D. P. Johns, and M. P. Swanney, "Spirometer users' and buyers' guide," Melbourne: National Asthma Council Australia, 2015.

[11] T. J. Barreiro and I. Perillo, "An approach to interpreting spirometry," *American Family Physician*, vol. 69, no. 5, pp. 1107-1114, 2004.

[12] F. D. McCool, "Global physiology and pathophysiology of cough: ACCP evidence-based clinical practice guidelines," *Chest*, vol. 129, no. 1, pp. 48S-53S, 2006.

[13] A. B. Chang, "The physiology of cough," *Paediatric Respiratory Reviews*, vol. 7, no. 1, pp. 2-8, 2006.

[14] W. D. Bennett and K. L. Zeman, "Effect of enhanced supramaximal flows on cough clearance," *Journal of Applied Physiology*, vol. 77, no. 4, pp. 1577-1583, 1994.

[15] K. Kosasih, U. R. Abeyratne, V. Swarnkar, and R. Triasih, "Wavelet augmented cough analysis for rapid childhood pneumonia diagnosis," *IEEE Transactions on Biomedical*

Engineering, vol. 62, no. 4, pp. 1185-1194, 2015.

[16] C. W. Thorpe, L. J. Toop, and K. P. Dawson, "Towards a quantitative description of asthmatic cough sounds," *European Respiratory Journal*, vol. 5, no. 6, pp. 685-692, 1992.

[17] M. V. A. Rao, N. K. Kausthubha, S. Yadav, D. Gope, U. M. Krishnaswamy, and P. K. Ghosh, "Automatic prediction of spirometry readings from cough and wheeze for monitoring of asthma severity," in *25th European Signal Processing Conference (EUSIPCO)*, Greece, 2017, pp. 41-45.

[18] G. L. Hall, B. R. Thompson, S. Stanojevic, M. J. Abramson, R. Beasley, A. Coates, *et al.*, "The Global Lung Initiative 2012 reference values reflect contemporary Australasian spirometry," *Respirology*, vol. 17, no. 7, pp. 1150-1, 2012.

[19] U. R. Abeyratne, V. Swarnkar, A. Setyati, and R. Triasih, "Cough sound analysis can rapidly diagnose childhood pneumonia," *Annals of Biomedical Engineering*, vol. 41, no. 11, pp. 2448-2462, Nov 2013.

[20] J. S. Reynolds, W. T. Goldsmith, J. B. Day, A. A. Abaza, A. M. Mahmoud, A. A. Afshari, *et al.*, "Classification of voluntary cough airflow patterns for prediction of abnormal spirometry," *IEEE Journal of Biomedical and Health Informatics*, vol. 20, no. 3, pp. 963-969, 2016.

[21] Y. A. Amrulloh, U. R. Abeyratne, V. Swarnkar, R. Triasih, and A. Setyati, "Automatic cough segmentation from non-contact sound recordings in pediatric wards," *Biomedical Signal Processing and Control*, vol. 21, pp. 126-136, 2015.

[22] X. Yan and X. G. Su, *Linear regression analysis: Theory and computing*. World Scientific Publishing Co., 2009.

[23] H. Drucker, C. J. C. Burges, L. Kaufman, A. J. Smola, and V. Vapnik, "Support vector regression machines," in *Advances in Neural Information Processing Systems*, 1997, pp. 155-161.

[24] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273-297, 1995.

[25] J. Snoek, H. Larochelle, and R. P. Adams, "Practical bayesian optimization of machine learning algorithms," in *Advances in Neural Information Processing Systems*, 2012, pp. 2951-2959.

[26] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504-507, 2006.

[27] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Networks*, vol. 61, pp. 85-117, 2015.

[28] D. E. Hinkle, W. Wiersma, and S. G. Jurs, *Applied statistics for the behavioral sciences*. Houghton Mifflin, 2003.

[29] J. Vestbo, S. S. Hurd, A. G. Agustí, P. W. Jones, C. Vogelmeier, A. Anzueto, *et al.*, "Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease," *American Journal of Respiratory and Critical Care Medicine*, vol. 187, no. 4, pp. 347-365, 2013.

[30] S. K. Singhi and H. Liu, "Feature subset selection bias for classification learning," in *Proceedings of the 23rd International Conference on Machine Learning*, Pittsburgh, Pennsylvania, USA, 2006, pp. 849-856.