

Classification Models for Pulmonary Function using Motion Analysis from Phone Sensors

Qian Cheng, MS^{1ad}, Joshua Juen, MS^{1bd},
Shashi Bellam, MD^{2a}, Nicholas Fulara, MS^{2b} Deanna Close, RN^{2b},
Jonathan C. Silverstein, MD^{2c}, Bruce Schatz, PhD^{1acd}

¹University of Illinois at Urbana-Champaign, Urbana, Illinois USA

^aDepartment of Computer Science, ^bDepartment of Electrical & Computer Engineering,

^cDepartment of Medical Information Science, ^dInstitute for Genomic Biology

²NorthShore University HealthSystem, Evanston, Illinois USA

^aDepartment of Medicine, ^bDepartment of Respiratory Therapy,

^cCenter for Biomedical Research Informatics

Abstract

Smartphones are ubiquitous, but it is unknown what physiological functions can be monitored at clinical quality. Pulmonary function is a standard measure of health status for cardiopulmonary patients. We have shown phone sensors can accurately measure walking patterns. Here we show that improved classification models can accurately measure pulmonary function, with sole inputs being sensor data from carried phones. Twenty-four cardiopulmonary patients performed six minute walk tests in pulmonary rehabilitation at a regional hospital. They carried smartphones running custom software recording phone motion. For every patient, every ten-second interval was correctly computed. The trained model perfectly computed the GOLD level 1/2/3, which is a standard categorization of pulmonary function as measured by spirometry. These results are encouraging towards field trials with passive monitors always running in the background. We expect patients can simply carry their phones during daily living, while supporting automatic computation of pulmonary function for health monitoring.

Keywords: *knowledge representation and information modeling, mobile health (patients), chronic care management (clinicians).*

Introduction

A revolution in health monitoring is coming, due to widespread mobile devices. Individual measurement can generate population cohorts of similar patients with similar status, so treatments can be effectively and efficiently targeted towards all groups¹.

Mobile phones are nearly ubiquitous in the United States, with the Pew Internet Project showing 91% ownership in May 2013, including 56% with smartphones. Even seniors over 65 years of age have 76% penetration of mobile phones². Since hundreds of millions of patients are already carrying phones, the opportunity appears for passive monitoring without adherence difficulties. We seek clinically valid physiological measures, vital signs which can be accurately monitored with smart phones.

Of the many measures that could be measured, the most important for diagnostic purposes are functional status. This is particularly true of cardiopulmonary disease, the major cause for chronic conditions in senior patients. As the patients age, their hearts and lungs slow down, and are not able to keep up with the demands. The physiological tests and measures clearly demonstrate their response to activity demand, e.g. as a patient walks, they slow down or move unsteadily when their heart and lungs cannot provide sufficient oxygen delivery to match increased exertion.

Pulmonary function is measured with a medical device called a spirometer. The patient breathes into this device, and the amount of air respired is recorded. The volume of air exhaled has been calibrated to provide standard measures of pulmonary function as discussed below. We show that a classification model can perfectly compute pulmonary function. In testing with cardiopulmonary patients, we further show that adequate inputs are the motion sensors already contained in ordinary smartphones. Simply carrying phones in daily living can measure health status.

For chronic heart and lung conditions, walk tests are widely used to assess the severity of the disease, including measures with accelerometer sensors^{3,4}. The Six-Minute Walk Test (6MWT) is a standard assessment⁵ for Chronic Obstructive Pulmonary Disease (COPD) and Congestive Heart Failure (CHF), which affects tens of millions of patients. A 6MWT measures the distance walked in six minutes back and forth over a fixed length walk way.

Normal gait requires many systems, including strength, sensation, and coordination, function in an integrated fashion, so abnormal gait is a diagnostic of many conditions⁶. Note that gait is the total walking pattern, the complete body motion, including swaying as well as stepping. So effective motion analysis must include more features than merely step counting, with adequate models.

Previously, we used the 6MWT to show motion features can be measured with smart phones carried by chronic patients, as sole input to a trained model that accurately predicts computes gait speed⁷ and oxygen saturation⁸, among other physiological measures for health status of chronic conditions. This study extends these results to computing pulmonary function, based on characteristic motions of health status. We also developed a phone app called MoveSense, and showed that it can record walking using phone sensors with similar accuracy to medical devices and higher accuracy than fitness devices⁹. This is true while running on the least expensive smart phones, such as the LG Optimus Zone, which now cost less than the portable fitness devices, such as the FitBit Flex.

Subjects and Data Collection

We recruited twenty-seven pulmonary patients at NorthShore University Health System, under IRB approval start November 2014. All chronic patients going through pulmonary rehabilitation in Respiratory Therapy in Evanston Hospital are now offered the option of participating in our study. Such patients are given our provided smartphones for recording their motion during a standard six-minute walk test during each visit, carried in a fanny pack. The six-minute walk tests are performed on a thirty-meter straight walkway in hospital corridor, with cones at the terminals. The patient walks back and forth on the walkway under the supervision of nurses for six minutes and the distance is recorded. The patient is permitted to stop and rest anytime, although the clock keeps going. Our software does not include this stationary data. All six-minute walk tests follow the ATS guidelines in a completely standard way⁵.

Pulmonary function tests are performed with a spirometer in clinical conditions¹⁰. In obstructive diseases, such as COPD, the ratio of forced expiratory volume in one second (FEV1) as compared to an age, gender, race, and height adjusted expected value is used as a sufficient indicator to measure severity level of the disease, called predicted FEV1%. The Global Initiative for Chronic Obstructive Lung Disease (GOLD) defines a standard for cardiopulmonary severity levels based on the predicted FEV1% values¹¹. There are four GOLD stages: GOLD 1 (mild), GOLD 2 (moderate), GOLD 3 (severe) and GOLD 4 (more severe). In this study, we have two GOLD 1 patients (predicted FEV1% ≥ 80), thirteen GOLD 2 patients (predicted FEV1%: 50-79) and nine GOLD 3 patients (predicted FEV1%: 30-49). GOLD 4 patients (predicted FEV1% < 30) typically cannot complete walk tests.

The demographic information for each group is shown in Table 1. There were two patients, Patients 12 and 24, who did not have pulmonary function tests in the record, so we eliminated these in our dataset. Patient 2 did not perform a six-minute walk test, which means there is no eligible walk data collected for this patient. This leaves twenty-four patients with walking data eligible for analysis with the models.

Table 1. Demographic information of each group of patients by GOLD levels. Age, height and weight are in average (minimum - maximum) format. These demographics plus gender are used to adjust the model.

	GOLD1	GOLD2	GOLD3	Overall
Number of Patients (Female)	2 (1)	13 (5)	9 (3)	24 (9)
Age [year]	69 (65-73)	80(67-95)	72 (55-85)	76 (55-95)
Height [m]	1.68 (1.55-1.80)	1.66 (1.24-1.83)	1.69 (1.55-1.83)	1.67 (1.24-1.83)
Weight [kg]	109.5 (93.0-126.1)	80.3(54.4-112.0)	81.2 (45.4-118.4)	83.0 (45.4-126.1)

Our phone app MoveSense was installed on the smartphones (a high-end Samsung Galaxy S5 and a low-end LG Optimus Zone2) for motion data collection. Our previous hardware experiment showed that the accelerometer in high-end and low-end smartphones are identical for human motion capturing, and both of them have comparable high quality as a medical accelerometer⁹. We take all data collected with the high-end smartphone for analysis and keep the low-end smartphone data for backup. Due to technical failure of the high-end phone, we use the data from low-end phone for three 6MWT sessions: the second session of Patient 5 and the first and the second session of Patient 16. We recorded the approximate starting time of each 6MWT session, and we let the patients stand still for seconds right before and after the 6MWT so that a pure six-minute walk can be easily extracted. The phone sensor data is downloaded into an archival database and correlated with patient medical data from the Epic EMR system.

Methodology

We established a model for GOLD stage classification. The model input is the raw three-dimensional acceleration collected by carried smartphones, and the output is a class of GOLD stage from 1 to 4 for each patient.

Preprocessing

MoveSense collects and processes the raw sensor data to extract qualified walking data. In six-minute walk test, subjects are allowed to rest during the walk test, so we must clean the dataset to detect if the patient is walking or stationary mode. We introduce a classification algorithm based on standard deviation analysis¹².

We calculate the magnitude of the raw acceleration data to compress 3D signals into 1D signals, then split the signal of each session into ten-second length samples with five-second sliding window⁷. Rather than a static threshold of standard deviation, the classification algorithm dynamically assigns standard deviation threshold for each session by reading the whole signal thread beforehand in one-second intervals. The algorithm can detect walking or stationary mode for six-minute walk test, see Figure 1. When the standard deviation threshold is set, each one-second walking has a binary decision for walking/not-walking. Then for all ten-second samples, if the ratio of walking is larger than 0.7, we select this sample as an eligible sample based on training sets.

Feature Selection

After preprocessing, we have obtained good walking data for each patient. We compute input features for training the model by feature selection approach (FSA)⁷. Feature selection approach (FSA) is based on empirical knowledge. A primary parameter to measure gait is the cadence, which is the number of strides within a unit of time¹³. We calculated cadence by counting steps in each ten-second interval¹⁴ and dividing by length of sample.

Related studies in motion tracking by wearable devices extract a series of features from raw acceleration data^{15,16,17}. We selected eight sufficient spatio-temporal gait parameters in both time and frequency domain. In the time domain, we selected mean and standard deviation of acceleration to describe the general distribution of the acceleration sample. In addition, mean crossing rate (MCR), root mean square (RMS), autocorrelation coefficient (AC) and coefficient of variance (CV) were calculated from time-series acceleration data. The mean crossing rate (MCR) represents the ratio of above and below acceleration. The root mean square (RMS) is a statistical measure on the variation of signal magnitude. The autocorrelation coefficient (AC) measures periodical similarity in time domain. The coefficient of variance (CV) is a normalized measure for dispersion of discrete samples. In the frequency domain, we computed the peak frequency (PF) and Shannon entropy. The peak frequency (PF) represents the frequency of peak magnitude in spectrum. Shannon entropy is an expected value of the information in the signal¹⁸.

Besides the cadence and spatio-temporal gait parameters, demographic information must also be considered in model training, just as a spirometer uses it to adjust the raw values from patient respiration. Unlike selecting demographic cohort and training models by cohort in our previous research⁷, we keep the four basic demographic parameters: age and sex, height and weight as factors in the input feature vector. With such input feature vector, we train universal models which ideally can be applied on general population. Overall, for each ten-second sample, the input feature vector \vec{x} contains thirteen independent features -- covering stepping, moving, and demographics.

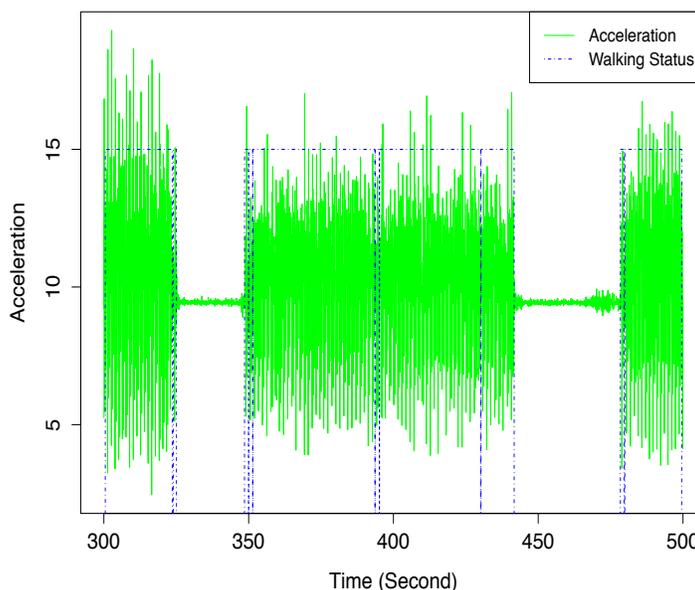


Figure 1. Sample result of walking detection algorithm. Green lines represent raw acceleration and blue dash lines represent walking status: higher level means this period is detected as walking and lower level means period is detected as stationary.

Model Training

Support vector machines (SVM) are a class of machine learning algorithms, widely used in learning classification models. We apply RBFSVM, the SVM with radial basis function kernel, to train the model for GOLD classification. SVM allows classification by remapping multi-dimensional vector \vec{x} into a higher dimensional space and determining a hyperplane or a set of hyperplanes that separate different classes¹⁹. The hyperplanes optimize the separation between data points in the given space, yielding the classification Equation (1). SVM classification requires a training set to find the variables α_i to determine $f(x)$ with a given kernel function and given inputs²⁰.

$$f(x) = \sum_{i=1}^N \alpha_i * K(x_i, x) \quad (1)$$

From Equation (1) we know that different kernel functions affect the overall performance of the trained model, previously we applied simple linear kernel in model training and obtain the technical baseline^{13,7}. In this study we applied RBFSVM, using the radial basis function kernel to train the SVM model²¹, as Equation (2). We set the hyper-parameter $\gamma = \frac{1}{D}$, based on the number of dimensions in training samples. \vec{x}_i and \vec{x}_j are two arbitrary sample vectors in the training set. In our dataset patient's status are labeled as one of the three GOLD stages (GOLD 1, GOLD 2 and GOLD 3). One-against-all strategy is applied to train this three-class classification model²². The tolerance of termination criterion is set as 0.001 and the insensitive-loss variable ξ is set as 0.1.

$$K(x_i, x_j) = \exp\left(-\gamma \left\|x_i - x_j\right\|^2\right), \quad \gamma = \frac{1}{D} \quad (2)$$

Model Validation

Since patients only take pulmonary function tests during the initial walk test (6MWT1) but not for the following test (6MWT2, 6MWT3), we only know the FEV1% predicted values for their first 6MWT session. We label samples during the first 6MWT with corresponding GOLD stages. Thus the dataset is partially labeled. We then select all labeled samples to form the training set. To avoid over-fitting, 10-fold cross validation is applied for self-validating the classification model²³. That is, randomly split the training set into ten parts, and each time select one out of the ten subsets as the test set and the other nine to train the model. After ten folds of model training and testing, we obtain predicted results of all samples in the dataset with a model trained by other samples from the same cohorts.

Voting Status with Thresholding

A classification model obtains GOLD stages for each ten-second walking samples. However, each six-minute walk test contains multiple ten-second walking samples, up to 37 if no stopping during the test. For cardiopulmonary patients, the severity will not change within a single walk test. But a walking sample may be affected by other factors and lead to false classification. Majority voting is robust rule to obtain single decision from multiple samples. Majority voting worked in our previous studies^{7,8} so we apply to obtain predicted GOLD stages for each patient. Based on empirical observations, we apply a strict threshold of 85% to accept the voting, which means only when the majority group gets more than 85% of the votes, the decision will be accepted. Otherwise, if none of the groups obtains higher than 85% votes, we mark the status as pending and clinically analyze the risk of this walking session.

Results

The dataset contains 1204 ten-second walking samples (39 six-minute walk tests) from 24 different patients, including 769 labeled samples from the first 6MWT session of each patient. We train classification models only using all labeled samples. Ten-fold cross validation is applied to self-validate the model. We then utilize the trained model to predict all unlabeled samples.

Evaluating the model

Besides the RBFSVM model, we train a classification model as in our previous study⁷, which applies SVM with linear kernel as baseline. The results are shown in Table 2. The overall accuracy of the RBFSVM model is 98.60% (100% for GOLD 1, 99.28% for GOLD 2 and 98.37% for GOLD 3). While the overall accuracy of the linear SVM model is 86.35% (100% for GOLD 1, 88.31% for GOLD 2 and 82.02% for GOLD 3). The RBFSVM model significantly improves classification accuracy for GOLD 2 and GOLD 3 samples from 82% to 98%.

Table 2. Self-validation results for RBF SVM model and linear SVM model. We compare the accuracy of classification, which is all correct classified samples divided by the total number of samples. This is by ten-second interval, not by patient. See Figure 2 for patient-level results.

	RBF SVM	Linear SVM
Overall	98.60%	86.35%
GOLD1	100%	100%
GOLD2	99.28%	88.31%
GOLD3	98.37%	82.02%

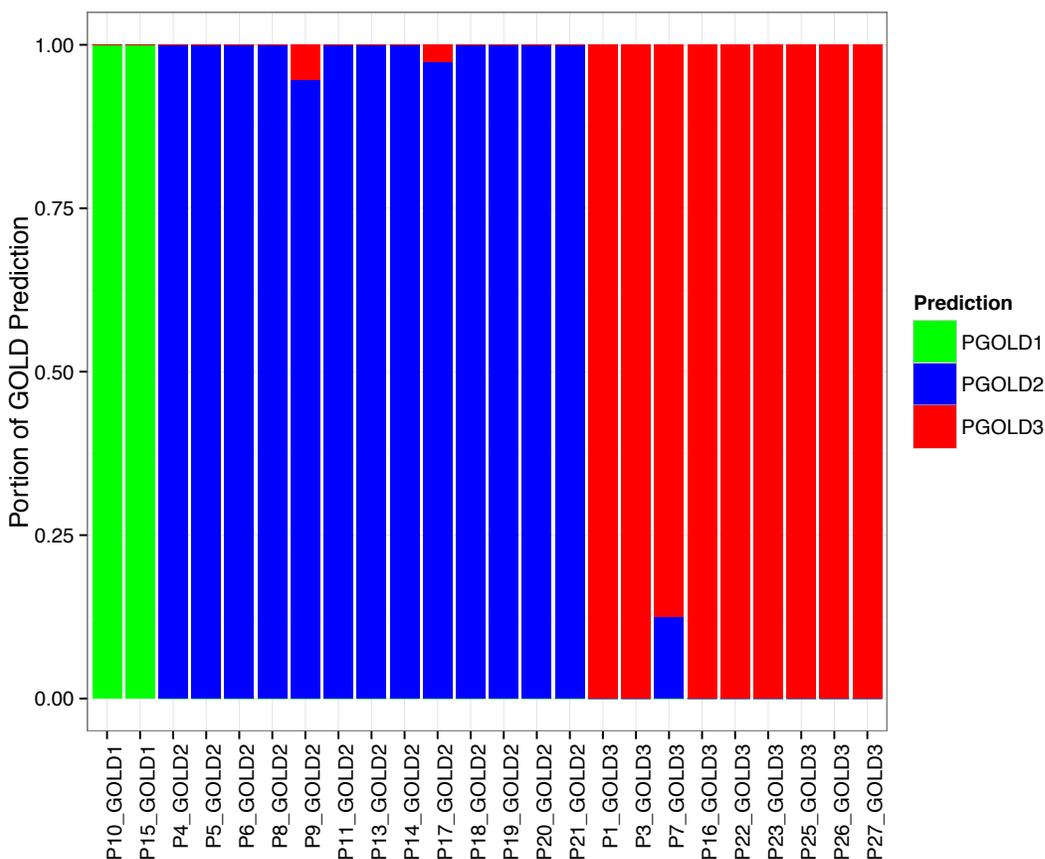


Figure 2. Evaluation of GOLD classification for labeled 6MWT sessions. The bar-plot shows the portion of predicted GOLD stages for all labeled walking samples of each patient. Except for Patient 7, 9 and 17, all other patients are 100% correctly predicted within every interval. After voting, all patients are predicted as their actual GOLD stages.

After obtaining the classification results for ten-second samples, majority voting is applied for categorizing each patient into a target GOLD stage. We perfectly categorize all patients into the corresponding GOLD stages summarized from their FEV1% predicted values, shown in Figure 2.

Classifying Unlabeled Samples

We train a universal classification model with all 769 labeled samples and apply this model to classify all unlabeled samples. There are 435 unlabeled samples from thirteen unique 6MWT sessions from ten patients, where three patients have third 6MWT sessions (5,16,21). Two out of the thirteen walking sessions are classified differently

from the patients' first labeled 6MWT session. For Patient 1, the second 6MWT session is 100% predicted as GOLD 2 while the first 6MWT shows the patient is in GOLD 3, as the FEV1 indicates with spirometry. For Patient 16, the second 6MWT session is 68.97% predicted as GOLD 2 and 31.03% predicted as GOLD 3, while the first 6MWT session of this patient is labeled as GOLD3, as the FEV1 indicates with spirometry. Referring to the rule of voting with thresholding, none of the predicted GOLD states reaches the 85% threshold so the second session of this patient is in pending status. Patient 16 also has a third 6MWT, which is 10.34% predicted as GOLD 1, 3.45% predicted as GOLD 2 and 86.21% predicted as GOLD 3. So the third session of Patient 16 is decided as GOLD 3, the same as the first session. For all other patients, the classifications of the unlabeled 6MWT sessions are 100% identical to their first 6MWT sessions for all intervals. The GOLD classification for each 6MWT session is shown in Figure 3.

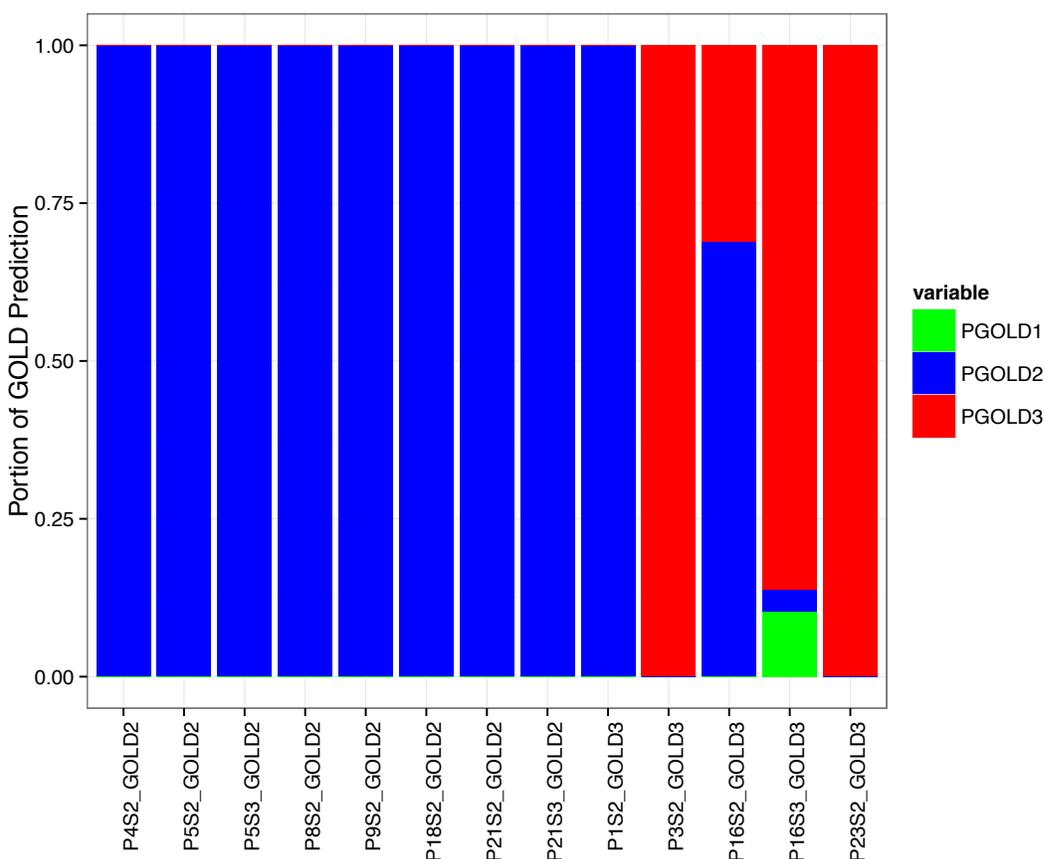


Figure 3. Evaluation of GOLD classification for unlabeled 6MWT sessions. The bar-plot shows the portion of predicted GOLD stages (PGOLD 1, PGOLD 2, PGOLD 3) for all unlabeled walking samples of each patient. Except for Patient 1 Session 2, Patient 16 Session 2 and Session 3, all other patients are 100% predicted as the GOLD stages they are labeled by the initial 6MWT session (Session 1).

Sufficient Walk Length for Accurate Classification

The main results above show that the classification model yields high accuracy when the model is trained with full six-minute walk for each patient session. Additionally, we analyze the effect of walk length in training progress to classification accuracy of the model. We randomly select different lengths of walking in the 6MWT session for each patient, from 30 seconds to six minutes (30s, 40s, 50s, 1min, 2min, 4min, 6min). We train and evaluate the model on the selected walking periods using the same strategy as above to obtain the predicting accuracy for each length. The result is shown in Figure 4. When the walk length is 120 seconds (two minutes), the predicting accuracy has reached 100% and remains there through the full six minutes. Note it is nearly perfect after just one-minute. So we assume a two-minute good walking for each patient is sufficient to train a classification model for detecting GOLD status. Here “good walking” refers to the steady pattern, as during a walk test, where high accuracy of the model is achieved, since the patient changes speeds reflecting their pulmonary function.

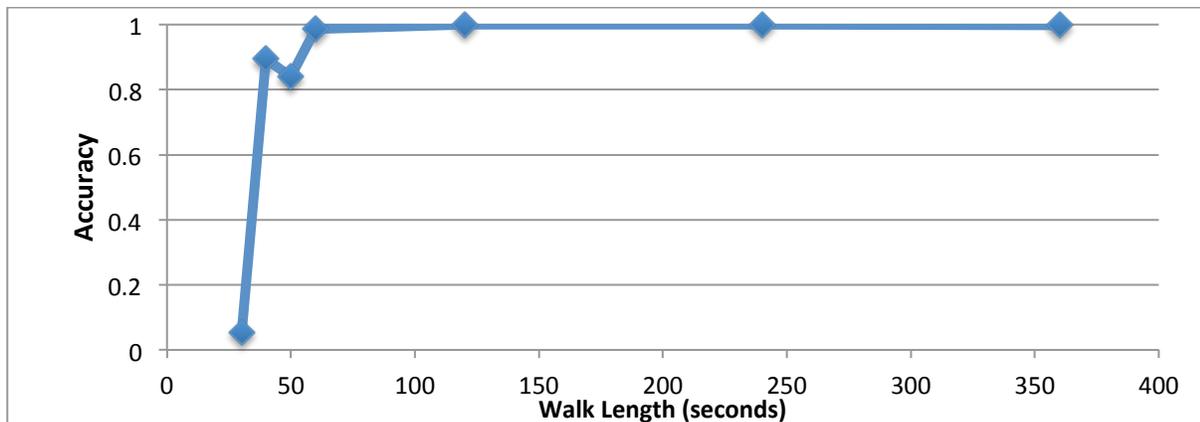


Figure 4. Classification Accuracy with Different Walk Length in Training. The accuracy in ten-fold cross validation reaches ~100% after two-minute (120-second) length.

Discussion

High Accuracy Classification on Labeled Walking Sessions

There were twenty-four patients participating in at least the first six-minute walk test session. The model does perfect computation for their pulmonary function. That is all patients are categorized into the correct GOLD stage. In this case the pulmonary function tests are performed the same day as the six-minute walk test, so we assume all patients' predicted FEV1% represents their current severity of respiratory limitation when we collect their motion data and the phone motion could fairly match the lung function. Figure 2 summarizes this perfect computation of health status for pulmonary function.

In more detail, for twenty-one patients, all their ten-second walking samples are identically predicted, and correctly reflect the patient's GOLD stage. This indicates spatio-temporal motion reflects more stable information than the one-dimensional walking speed. Patient may speed up or slow down during the six-minute walk test, but the spatio-temporal motion will always yield the correct health status, by measuring stability deviation of walking movements. Based on this outcome, we assume that six-minute walk tests are not necessary for detecting health status of cardiopulmonary patients. Instead, if we can find any good walking sessions that are identical to the walking during the six-minute walk test, spatio-temporal motion data from such walking session can be used to compute GOLD status. As Figure 4 shows, our model only needs 2 minutes of good walking for perfect computation.

For three patients, not all ten-second samples are classified correctly, including Patient 7, Patient 9 and Patient 17. For Patient 9, one ten-second sample out of thirty-three eligible samples gets false classification, while for Patient 17, one ten-second sample out of twenty-five eligible samples gets false classification. These can be simple statistical error so we can ignore them. For Patient 7, five samples out of thirty-seven eligible samples get false classification, which takes 13.5% of the overall classification classifying GOLD 2 for a patient measured as GOLD 3. So voting with 85% threshold would still predict correct status compared to spirometry that measures FEV1%.

In addition, for Patient 7, the spirometry measurement was 5 months earlier than the 6MWT, whereas nearly all other patients had their PFT within 1 month. This delay could lead to false classification, due to change in clinical status. For this patient, they also had a medication change to begin COPD inhalers during this period, so we expect an improved status with better pulmonary function. Thus there is a clinical explanation for lower model inaccuracy.

Identical Classification on Unlabeled Following Walking Sessions

From medical literature, six-minute walk distance (6MWD) is considered as a gold standard for diagnosis of COPD severity and other cardiopulmonary health status, but it is more accurate for severe COPD than moderate or mild COPD²⁴. In our experiment, ten patients performed at least two six-minute walk test sessions, as shown in Figure 3, usually spaced several weeks apart. We compare their initial 6MWDs and following 6MWDs, shown in the Bland-Altman plot in Figure 5.

There are 13 sessions, since 3 patients also did a session 3, but only 11 points since we eliminate 2 because no actual 6MWD recorded (Patient 5 Session 3 and Patient 16 Session 3). The initial 6MWDs are in average of 310.8m ($\pm 95.16m$) and the following 6MWDs are in average of 367.1m ($\pm 106.5m$). Generally, the patients walked further (had greater 6MWD) in later sessions, as expected for rehabilitation patients.

Our analysis indicates that 6MWD is not sufficient for measuring COPD severity levels, certainly compared to pulmonary function for cardiopulmonary patients. In our dataset, the GOLD 2 and GOLD 3 patients are not well distinguished by 6MWD. The 6MWD of GOLD 3 patients can be as high as 350 meters but the 6MWD of GOLD 2 patients can be below 250 meters. Note the variation between two different 6MWDs of a patient becomes larger when the distance is shorter, e.g. the second 6MWD of one GOLD 3 patient (Patient 23) is 150 meters longer than the initial 6MWD, even though performed just twenty days after he began the rehabilitation process.

Our classification model again had high accuracy, since such variation on total distance does not affect the stability of patients' motion pattern. Eight out of ten patients are detected strictly as the same GOLD status for their second or third 6MWT sessions. All of the ten-second walking samples of their following 6MWT sessions are identical to their original GOLD status, even for Patient 23, the one with the largest variation of 6MWDs. This "unfair" 8 of 10 model accuracy for later sessions compared to the initial GOLD status is shown in Figure 3.

The two patients who are not correctly classified at subsequent sessions are Patient 1 and Patient 16. For Patient 1, GOLD 2 is classified at the subsequent session but the PFT measure indicates GOLD 3. Patient 1 had a clinical diagnosis of both COPD and CHF when he was enrolled in cardiac rehabilitation. His medication regimen for CHF was adjusted at the start of his rehabilitation program and his motion was noted to improve, likely from better control of his CHF, so in fact this "unfair" comparison of initial GOLD status to improved GOLD status with the model may be measuring actual improvement rather than model inaccuracy. Congestive Heart Failure is less closely correlated to pulmonary function than Chronic Obstructive Pulmonary Disease, limiting the usefulness of GOLD criteria for defining the severity of functional limitation in CHF. The second 6MWD of Patient 1 is 72.5 meters longer than the initial 6MWD.

Monitoring Severity Change Longitudinally

Patient 16 had diagnosis of sarcoidosis, which is a lung disease that has variable effects on pulmonary function. Sarcoidosis cannot be well categorized by disease severity based upon GOLD criteria, as the physiologic limitations are less clearly correlated with predicted FEV1%, as in COPD. However, this patient was measured for pulmonary function via spirometry for FEV1 indicating GOLD 3, which was correctly classified in Session 1.

Patient 16 performed three 6MWT sessions at different visits, among which the second session was about one month later than the initial session and the third session is about five month away from the initial session, which is beyond the three-month rehab period. In Session 2, there was mixed classification of 2/3 intervals indicating GOLD 2 and 1/3 indicating GOLD 3. The patient was likely within transition between status levels, due to rehabilitation progress.

During Session 3, not only was it much later, but there was a clinical note that the patient had suffered a significant illness which was the reason to return to rehabilitation. This session was the only true mixture, with some classifications of GOLD 1/2/3, although 86% of the intervals indicated GOLD 3 once again. So the lung

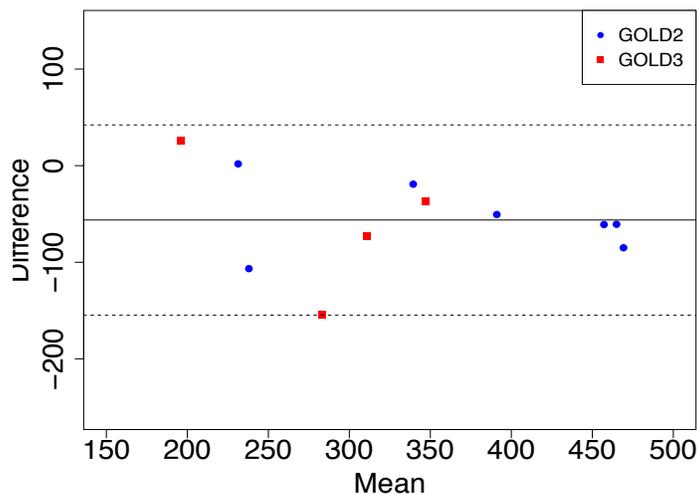


Figure 5. Bland-Altman plot of 6MWD comparison. Blue dots represent 6MWDs GOLD 2 patient sessions. Red squares represent 6MWDs of GOLD 3 patient sessions. GOLD severity levels are not well separated by 6MWD.

physiological function likely worsened and the patient regressed back to the original status level. Our model would have predicted this correctly, if the threshold mentioned earlier of 85% was implemented. Thus there is clinical evidence that our model can predict changes in pulmonary function.

Conclusion

In this study, we expand our previous results⁷ into accurate computation of pulmonary function with universal models. This requires utilizing demographic features with better training sets (ATS standard) and better statistical models (RBF-SVM), which utilize characteristic motions in addition to step counting. Every patient now has correct modeling of GOLD level, even sample by sample (10 seconds) not only walktest by walktest (6 minutes). Measuring motion is a potential solution for passive monitoring, which distinguishes this work from the phone applications measuring pulmonary functions with the microphone²⁵. A passive monitor has compliance advantages over active phone applications. That is, the patient simply uses their personal phone as usual during daily activities, no special actions or special experts are necessary. The microphone “spirometer” also has adherence difficulty, the patient must breathe properly into the proper place, also a difficulty with a medical spirometer.

The trained model is providing perfect computation of pulmonary function category (classifying GOLD stage). This is especially true for the first session with pulmonary patients, where spirometry has just measured the pulmonary function so the pulmonary function value can be directly compared to the motion value. For senior patients undergoing pulmonary rehabilitation, respiration and motion are closely correlated. This is why the classification is so accurate, especially in the hospital setting with monitored walk test. The result is encouraging, so we shall be relaxing environmental constraints on our measurements in forthcoming studies, moving closer to daily activities in the real world of cardiopulmonary patients. Under expanded IRB, we now allow selected patients to bring smart phones home from rehab, to measure their motions and predict their pulmonary functions during daily activities.

The home trials are more complex than the hospital trials thus far. The patient will be doing many different activities throughout the day, while the periods when they are carrying their phones could all be potentially recorded by a passive monitor that is always running. The passive monitor phone app used in the home is a re-implementation of the active monitor used in the clinic, which only records when the phone is in steady motion and only archives when the patient is good walking. Since the models have high accuracy during walk tests, we are developing software for activity recognition. This is a well-studied problem in computer science, for using motion sensors to detect which activity a person is performing^{17,26}. General activity recognition is hard, but our problem is a specialized version of that. The recognition filter in the phone app need only detect good walking, when the body motion is similar to a walk test. This paper reports on initial progress of autodetecting stopping during walktests. Since the amount of walking needed for accurate computation is very small, only 1 or 2 minutes per session, the filter can be tuned up, for high precision only during periods when there is definite measurement of good walking.

We are optimistic that full-scale clinical trials are possible to compute health status for cardiopulmonary patients. They need only carry their own phones and do some limited good walking during the day when the passive monitor is recording unobtrusively. Since the cheapest smartphones can accurately support this computation, large-scale population measurement can become an everyday reality for health systems.

Acknowledgements

At the University of Illinois at Urbana-Champaign, the Institute for Genomic Biology provided facilities for software development and data analysis. At NorthShore University HealthSystem, the Department of Respiratory Therapy at Evanston Hospital carried out patient testing and phone recording, with co-author Deanna Close, RN, as supervising pulmonary nurse. All clinical experiments were protected by NorthShore IRB EH15-025 entitled “Mobile phone software MoveSense versus traditional exercise testing to assess health status”, with co-author Shashi Bellam, MD, as supervising pulmonary physician. We acknowledge Brian Edwards and Catherine Zhu for managing and anonymizing phone data and medical records inside NorthShore University HealthSystem using the REDCap archiving system²⁷.

References

1. Schatz B and Berlin R. Healthcare Infrastructure: Health Systems for Individuals and Populations. Series in Health Informatics. London: Springer Limited, 2011.
2. Fox S and Duggan M. Tracking for health. Pew Research Internet Project, Jan 2013.
3. Jehn M, Schmidt-Trucksäess A, Schuster T, et al. Accelerometer-based quantification of 6-minute walk test performance in patients with chronic heart failure: applicability in telemedicine. *J Cardiac Failure*, 2009;15(4):334-340.
4. Annegarn J, Spruit M, Savelberg H, et al. Differences in walking pattern during 6-min walk test between patients with COPD and healthy subjects. *PLOS ONE* 2012;7(5):e37329.
5. ATS statement: Guidelines for the six-minute walk test. *Amer J Respiratory Critical Care Med* 2002;166(9):111-117.
6. Longo D, Fauci A, Kasper D, Hauser S, Jameson J, Loscalzo J. Part 17: Neurologic disorders. In: Harrison's Principles of Internal Medicine. 18th ed. New York: McGraw-Hill, 2011.
7. Juen J, Cheng Q, Prieto-Centurion V, Krishnan JA, Schatz B. Health monitors for chronic disease by gait analysis with mobile phones. *Telemedicine and e-Health* 2014;20(11):1035-1041.
8. Cheng Q, Juen J, Hsu-Lumetta J, Schatz B. Predicting Transitions in Oxygen Saturation using phone sensors. *Telemedicine and e-Health* 2016;22(2):132-137.
9. Juen J, Cheng Q, Schatz B. A natural walking monitor for pulmonary patients using simple smart phones. *IEEE J Biomedical and Health Informatics* 2015;19(4):1399-1405.
10. Broaddus V, Mason R, Ernst J et al. Murray & Nadel's Textbook of respiratory medicine. Elsevier Health Sciences, 2015.
11. Rabe K, Hurd S, Anzueto A, et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. *Amer J Respiratory Critical Care Medicine* 2007;176(6):532-555.
12. Wallace C and Boulton D. An information measure for classification. *The Computer Journal*, Br Computer Soc, 1968;11:185-194.
13. Whittle M. Gait analysis: an introduction. Butterworth-Heinemann, Fourth edition, 2014.
14. Cheng Q, Juen J, Li Y, Prieto-Centurion V, Krishnan J, Schatz B. GaitTrack: Health Monitoring of Body Motion from Spatio-Temporal Parameters of Simple Smart Phones. *ACM BCB International Conference on Bioinformatics, Computational Biology and Biomedical Informatics*, 2013;897.
15. Zijlstra W and Hof A. Assessment of spatio-temporal gait parameters from trunk accelerations during human walking. *Gait and Posture* 2003;18(2):1-10.
16. Nishiguchi S, Yamada M, Nagai K, et al. Reliability and validity of gait analysis by Android based smartphone. *Telemedicine and e-Health* 2012;18(4):292-296.
17. Lu H, Yang J, and Liu Z, et al. The Jigsaw continuous sensing engine for mobile phone applications. *ACM SENSYS Embedded Networked Sensor Systems*. 2010.
18. Shannon C. A mathematical theory of communication. *ACM SIGMOBILE Mobile Computing and Communications Review*, ACM, 2001;5:3-55.
19. Cortes C and Vapnik V. Support-vector networks. *Machine learning*, Springer, 1995;20:273-297.
20. Chang C and Lin C. LIBSVM: a library for support vector machines. *ACM Trans Intelligent Systems and Technology* 2011;2(3):27.
21. Keerthi S and Lin C. Asymptotic behaviors of support vector machines with Gaussian kernel. *Neural computation*, MIT Press, 2003;15:1667-1689.
22. Hsu C and Lin C. A comparison of methods for multiclass support vector machines. *IEEE Transactions on Neural Networks*, 2002;13:415-425.
23. Kohavi R et al. A study of cross-validation and bootstrap for accuracy estimation and model selection. *International Joint Conference on Artificial Intelligence (IJCAI)*, 1995;14:1137-1145.
24. Pinto-Plata V, Cote C, Cabral H et al. The 6-min walk distance: change over time and value as a predictor of survival in severe COPD. *European Respiratory Journal*, Eur Respiratory Soc, 2004;23:28-33.
25. Larson E, Goel M, Boriello G, Heltshe S, Rosenfeld M, Patel S. SpiroSmart: using a microphone to measure lung function on a mobile phone. *ACM Int Conference on Ubiquitous Computing*, 2012: 280-289.
26. Kwapisz J, Weiss G, Moore S. Activity recognition using cell phone accelerometers. *ACM SIGKDD Explorations Newsletter*, 2011;12:74-82.
27. Harris P, Taylor R, Thielke R, et al. Research electronic data capture (REDCap) – A metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomedical Informatics*, 2009;2(2):377-381.