Published in final edited form as:

ACM BCB. 2016 October ; 2016: 41–49. doi:10.1145/2975167.2975171.

Mining Discriminative Patterns to Predict Health Status for Cardiopulmonary Patients

Qian Cheng,

Computer Science, UIUC, Champaign, IL

Jingbo Shang,

Computer Science, UIUC, Champaign, IL

Joshua Juen,

Electrical and Computer, Engineering, UIUC, Champaign, IL

Jiawei Han, and

Computer Science, UIUC, Champaign, IL

Bruce Schatz

Medical Information Science, UIUC, Champaign, IL

Abstract

Smartphones are ubiquitous now, but it is still unclear what physiological functions they can monitor at clinical quality. Pulmonary function is a standard measure of health status for cardiopulmonary patients. We have shown that predictive models can accurately classify cardiopulmonary conditions from healthy status, as well as different severity levels within cardiopulmonary disease, the GOLD stages. Here we propose several universal models to monitor cardiopulmonary conditions, including DPClass, a novel learning approach we designed. We carefully prepare motion dataset covering status from GOLD 0 (healthy), GOLD 1 (mild), GOLD 2 (moderate), all the way to GOLD 3 (severe). Sixty-six subjects participate in this study. After de-identification, their walking data are applied to train the predictive models. The RBF-SVM model yields the highest accuracy while the DPClass model provides better interpretation of the model mechanisms. We not only provide promising solutions to monitor health status by simply carrying a smartphone, but also demonstrate how demographics influences predictive models of cardiopulmonary disease.

Keywords

Chronic Disease Assessment; Free-living Health Monitoring; Discriminative Pattern Mining

1. Introduction

A revolution in health care is coming, due to the widely spread and ubiquitous usage of mobile devices. Patients with chronic diseases now live for many years with decreasing function. Mobile devices have the potential to monitor the status change for chronic patients and help guide them to a more effective treatment plan. Individual measurement for chronic

disease should be adjusted with population information so the measurement can be applied towards all persons within the population ¹.

Smartphones are ubiquitous in the United States, and spreading rapidly in the rest of the world. The most recent Pew Internet Study shows 64% of adults own smartphones in the United States ². There is opportunity for passively monitoring chronic disease without adherence obstacles, in the near future, as large numbers of senior adults routinely utilize these. Thus we seek vital signs which can be accurately monitored by simple smartphones, as clinically valid physiological measures.

Chronic Obstructive Pulmonary Disease (COPD), a major cardiopulmonary disease affecting tens of millions of patients, is usually measured by pulmonary function tests (PFTs)³. Pulmonary function is measured with a spirometer in clinical condition. The patient breathes into the spirometer and the airflow is measured for such action. The forced expiratory volume in 1 second (FEV1) has been calibrated by forced vital capacity (FVC) and normalized by subject's demographic information to provide standard measures function, called predicted FEV1%. The Global Initiative for Chronic Obstructive Lung Disease (GOLD) defines four different stages (GOLD 1 to GOLD 4) representing mild, moderate, severe and more severe status of COPD. In this study we also define healthy subjects as "GOLD 0". Due to lack of mobility of GOLD 4 patients, our predictive models focus on classifying status from GOLD 0 (healthy) to GOLD 1 (mild) to GOLD 2 (moderate) to GOLD 3 (severe).

Measuring body motion with accelerometer sensors during a walk test is widely applied to assess the severity of chronic heart and lung diseases^{4, 5}. Six-minute walk test (6MWT) is a standard clinical measure for COPD and congestive heart failure (CHF)⁶. This test is a six-minute walk back and forth over a fixed length of walk way (usually thirty meters), with the clinical record showing the total distance (6MWD). Normal gait requires many systems, including strength, sensation and coordination, function in an integrated fashion, so abnormal gait detection is a diagnostic of many conditions ⁷. In general, gait is walking pattern, the complete body motion including swaying and stepping, more complicated than speed⁸. Thus, sufficient motion analyzer must include more information than simply counting steps or measuring distance, with adaptive predictive models. Our previous work indicates statistical learning models can be successfully applied to health status measures. However, a major problem related to the mechanism restricts the application of these models in medical domains, namely the lack of interpretation for the significance of input features. Decision trees provide an explicit way to interpret how a learning model works, but the deterministic manner of a tree structure limits the predicting power. We have developed a novel pattern mining approach, the DPClass⁹, yielding an optimized solution with higher predicting power and better interpretable mechanism.

We have developed an application, MoveSense¹⁰, for Android smartphones. It serve as a framework for recording walking, transmitting data and performing status prediction. In this study, we validate models trained with different supervised learning techniques. Meanwhile, we obtain more insights from tree-based models to better interpret the model prediction mechanisms.

2. Subjects and Data Collection

Our laboratory collaborated with two health systems in Illinois, the NorthShore University HealthSystem, a major health system located in urban Chicago area, and Carle Foundation Hospital, a major health system in rural Champaign-Urbana area. With IRB approval, totally sixty-six subjects participate in this study, including fifty-five pulmonary patients and eleven healthy subjects. All subjects performed at least a full session of standard six-minute walk test (6MWT), walking back and forth for six minutes on a straight walkway, under the supervision of nurses. Subjects are permitted to stop and rest anytime during the walk test, although the clock keeps going. Our software automatically eliminates the stationary part of the collected data. The distance of 6MWT was recorded for each subject. All six-minute walk tests follow the ATS guidelines ⁶.

Pulmonary function tests are performed with a spirometer in clinical conditions ³. 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%. Based on the predicted FEV1% values, the Global Initiative for Chronic Obstructive Lung Disease (GOLD) defines a standard measure for cardiopulmonary diseases, evaluating the severity levels from mild (GOLD 1, FEV1%: 80), moderate (GOLD 2 FEV1%: 50 - 79), severe (GOLD 3, FEV1%: 30 - 49) and more severe (GOLD 4, FEV1%: < 30). In this study, all healthy subjects are labeled as GOLD 0, indicating that they do not have cardiopulmonary diseases. Since the more severe (GOLD 4) patients have significant lack of mobility, measuring their health status is a difficult with motion sensors. So we only focus on GOLD 0 to GOLD 3 in this study. There are five cardiopulmonary patient missing PFT values so we eliminate them. Two healthy subjects do not have height and weight information so we eliminate them as well. Overall, there are fifty-nine subjects data: nine GOLD 0 subjects, twenty-three GOLD 1 subjects, eighteen GOLD 2 subjects and nine GOLD 3 subjects. The demographic information is shown in Table 1.

Our phone application MoveSense¹⁰ was installed on the smartphones (a high-end Samsung Galaxy S5 cost \$600 and a low-end LG Optimus Zone2 cost \$60) for motion data collection. Previous hardware experiment showed that the accelerometers in high-end and low-end smartphones are identical for human motion capturing, and both of them are equally performing as a high-quality medical accelerometer. We primarily collect data from high-end smartphone for analysis and keep the low-end smartphone data for backup. Once the high-end phone failed, we will use the low-end phone data instead since we previously showed that the low-end phone keeps equivalent quality of sensor data as high-end phones¹⁰. After de-identification, the phone sensor data, together with other demographic information and test results, are stored into an archival database.

3. Methodology

In this study we established four models for GOLD stage prediction. The models accept raw three-dimensional acceleration, collected by MoveSense software, as input. After

processing, the predictive models output a specific GOLD stage out of the defined four GOLD stages: GOLD 0, GOLD 1, GOLD 2 and GOLD 3.

3.1 Data Preparation

As described in our previous work, MoveSense process the raw acceleration and extract qualified walking data. This process is necessary because in standard 6MWT, subjects are allowed to rest anytime during the walk test and we must detect if the patient is steadily walking or stationary. We developed an algorithm for detecting walking based on standard deviation analysis ¹¹.

We calculate the magnitude of the raw acceleration to compress three-dimensional signals into one-dimensional signals. Each walk test session was split into ten-second samples with a five-second sliding window. The walking detection algorithm sets dynamic threshold of standard deviation by reading whole thread of signals in the walk test session beforehand. For each ten-second sample, if the walking portion is larger than 0.7, we keep this sample as a good walking sample, otherwise, we dismiss it.

3.2 Feature Selection

We obtain good walking data for each subject after preprocessing. Feature selection approach (FSA), which is based on empirical knowledge of medicine and kinesiology, is applied to compute the input features ¹². Cadence, walking strides per minute, is an essential parameter to measure gait speed. We compute cadence by counting steps in each ten-second interval ¹³.

Related studies extract a series of features from raw acceleration to compute gait features^{14, 15}. As in our previous study, we computed eight sufficient spatio-temporal gait parameters in both time and frequency domain. In time domain, mean and standard deviation of acceleration are computed to describe the general distribution of motion. In addition, several signal processing features describing time-series signals are computed: mean crossing rate (MCR), root mean square (RMS), autocorrelation coefficient (AC) and coefficient of variance (CV). In frequency domain, we computed peak frequency (PF) and Shannon entropy.

We integrate demographic information of subjects to complete the input vector. Our predictive models need basic demographic information for calibration, to effectively match the predicted spirometry values which also integrate these. Unlike selecting demographic cohort and training models by cohort in our previous study ¹⁶, we keep the four basic demographic parameters: age and sex, height and weight in the input vector. Ideally, we are capable to train universal models which can be applied to all population since demographic information has been involved. Overall, for each ten-second sample, the input feature vector $\vec{x}$ contains thirteen independent features – covering stepping, moving, and demographics.

3.3 Model Training

Support vector machines (SVM) are a class of machine learning algorithms, operating classification by remapping training data into a higher dimensional space and determining

sets of hyperplanes, also called “support vectors”¹⁷. An SVM model optimizes the separation between data points in the input feature space, yielding the classification Equation 1, which requires a training set to simply find the variables c_i to determine $f(x)$ given the kernel function and inputs.

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

$K(x_i, x_j)$ is the kernel function, which determines the optimization strategy. In our previous study we applied a Fisher's linear kernel^{18, 16}, defined as Equation 2.

$$K(x_i, x_j) = U_{x_i}^T U_{x_j} \quad (2)$$

In this study, we train linear SVM model as baseline to evaluate the performance of other predictive models. In order to improve the overall performance of SVM, we apply a different kernel function, the radial basis function, Equation 3. D is the number of dimensions in training samples. The tolerance of termination criterion is set as 0.001 and the insensitive-loss variable ϵ is set as 0.1.

$$K(x_i, x_j) = \exp\left(-\frac{1}{D} \|x_i - x_j\|^2\right) \quad (3)$$

Decision trees are a set of decision making strategies that can be represented as a tree structure. A decision tree model can be established in a supervised behavior with labeled training set. The Gini index, shown as Equation 4, is usually applied for tree model optimization.

$$\text{Gini}(D) = 1 - \sum_{i=1}^m p_i^2 \quad (4)$$

We applied recursive partitioning (*rpart*) routines¹⁹ to obtain classification tree for GOLD prediction. With default setting the *rpart* program will take the input data and label, and then build the tree model. It also helps plotting the general tree structure so that we can observe details how the tree model is making classification decisions.

In this study, we applied DPClass⁹, an effective but concise discriminative pattern-based classification framework, to obtain a predictive model for GOLD, where a discriminative pattern is a conjunctive clause of several conditions, for example, (Age < 80.5) AND (Height > 1.5). DPClass can automatically extract the top discriminative patterns in model

training. The generated discriminative patterns, or discriminative rules reflect interactions of all input features.

Pattern-based classification, such as PatClass and DDPMine, was originally proposed to improve the prediction accuracy by constructing high-order interactions among features using hundreds or thousands of frequent patterns. Multi-tree based models, such as Random Forest, have demonstrated advantages of classification accuracy and handling both numerical and categorical features. DPClass takes the advantages of both concepts and proposes to resolve classification problems using discriminative patterns.

Candidate Discriminative Patterns—As shown in Figure 1, DPClass first starts from training data and train many (e.g., 100) random decision trees²⁰. Every prefix path from the root of any tree is adopted as a candidate discriminative pattern. By the default setting, about 10 thousand candidate discriminative patterns are usually generated.

Pattern Space—Now, instances in their original feature space can be mapped into the pattern space using a set of discriminative patterns. For each discriminative pattern, there is one corresponding binary dimension describing whether the instances satisfy the pattern or not. Because the dimension of the pattern space is equal to the number of discriminative patterns which is a very large number after the generation phase, we need to further select a limited number of patterns and thus make the pattern space small and efficient.

Top-k Discriminative Patterns—L1 regularization (i.e., LASSO²¹) is designed to make the weight vector sparse by tuning a non-negative parameter λ . Since we are actually selecting features in the pattern space, for a given λ , we optimize the following loss function to get a subset of patterns which are most important.

$$\mathcal{L} = \sum_i^n l(\mathbf{x}'_i{}^T \mathbf{w}, y_i) + \lambda \cdot \|\mathbf{w}\|_1 \quad (5)$$

where, $\mathbf{x}'_i$ is the mapped binary feature representation in pattern space of i -th instance; y_i is the label of i -th instance; $\mathbf{w}$ is the weight vector in the generalized linear model; $l(\cdot, \cdot)$ is a general loss function such as logistic loss. To ensure there are at most k patterns having non-zero weights in the pattern space, the binary search technique can be utilized to automatically determine λ , assuming that there exists some hidden importance ranking among features and once the weight of a feature becomes non-zero in a given $\lambda = \nu$, it will also be non-zero for almost any smaller $\lambda < \nu$.

Final Multi-class Classification—After selecting top-30 discriminative patterns, we have created a 30-dimensional pattern space and feed these features to train a multi-class classification model, where we adopt the implementation in GLMNET²².

4. Results

The dataset contains 3,748 ten-second walking samples. Each sample is labeled as the GOLD stage of the corresponding subjects, from GOLD 0 to GOLD 3. We trained different supervised models for the same input with different learning methods we described above. We applied standard ten-fold cross validation for model validation to avoid over-fitting. Then, randomly split the training set into ten parts, and each time select one out of the ten folds as the test set and the other nine folds to train the models. After all ten folds being tested by models trained with other folds, we obtain the prediction results of all samples in the manner that none of these samples contributes to model training.

4.1 Predictive models accurately detect different GOLD stages

We trained four different predictive models for GOLD stages, a linear SVM model, a RBF-SVM model, a decision tree model and a DPClass model. The validation results are shown in Table 2. The linear SVM model predicts GOLD stages with 78.36% accuracy. The RBF-SVM model obtains 99.15% predicting accuracy, which is the highest among all tested models. The decision tree model returns 94.56% accuracy and DPClass returns 95.97% predicting accuracy. All four predictive models can distinguish healthy subjects and cardiopulmonary patients very well. However, distinguishing the slowest patients GOLD 2 and GOLD 3 is the most difficult task for all predictive models. The linear SVM model has very low classification power for distinguish GOLD 2 and GOLD 3. The confusion matrices are shown in Table 3 for all four models.

Each six-minute walk test contains multiple ten-second walking samples. For cardiopulmonary patients, the severity will not change within a single walk test, while other factors can affect a single walking sample so it may lead to false prediction. Majority voting strategy is applied to summarize a single GOLD stage outcome from all walking samples in the same walk test session.

Only the linear SVM model has false prediction on subject level. Two subjects are misclassified: CF015, a GOLD 1 to be predicted as GOLD 2, and CF003, a GOLD 2 to be predicted as GOLD 1. All other predictive models have perfect prediction on subject level. That is to say, all healthy and cardiopulmonary patients can be categorized to the correct cohort group they belong to.

4.2 Tree model returns dominant features for GOLD prediction

An advantage of tree structure models is that it outlines the classification strategy based on input features. We trained a universal model with all samples and obtains the classification structure. The misclassification error rate is 0.0267. The number of terminal nodes are 25. The minimum depth of the tree is one while maximum depth of the tree is six. Three demographic features are used to construct the tree: age, height and weight. Five spatio-temporal gait parameters are used to construct the tree: MeanAcc, StdAcc, RMS, PF and Shannon Entropy. All young subjects (age < 49.5) are classified as healthy. For GOLD detection, both demographics and gait features affect the results. The most crucial gait features for classification is the mean of acceleration (MeanAcc), standard deviation of

acceleration (StdAcc) and root mean square (RMS). A brief decision structure is shown in Figure 2

4.3 DPClass model returns top discriminative patterns for GOLD prediction

Mining discriminative patterns is the major novelty of DPClass⁹. Unlike decision trees, which directly make decisions on each node by a specific input feature, DPClass selects the top discriminative patterns and automatically generates decision rules from any combination input features. For this specific model, a sample decision rule could be “(Age<80.5) and (Weight>= 63.28) AND (MeanAcc >= 9.67) AND (StdAcc>= 1.09936) AND (RMS< 1.05517)”, which means if a sample satisfied this rule, the program will return a 1 in the generated feature space. Otherwise, the program will return 0. From the patients datasets, we generated the top 30 rules (discriminative patterns) for GOLD prediction, shown in Table 5. The self-evaluation accuracy of DPClass model is 0.9789.

Each discriminative pattern contains threshold for several input features. Unlike the decision tree model, where demographics functions as standalone nodes for decision making, in the DPClass model, demographics participated on every single rule. That is, there is at least one feature out of age, sex, height and weight threshold influencing the binary decision making for every of the top 30 rules. In more detail, age participates in 24/30 of the rules, sex participates in 2/30 of the rules, height participates in 12/30 of the rules and weight participates in 18/30 of the rules. Even though cadence is an important gait feature in kinesiology, it only affects two out of thirty rules in the DPClass model for motion correlating to function. Except for Shannon entropy, all spatio-temporal gait parameters contribute to the discriminative rules.

5. Discussion

We have studied monitoring health status for cardiopulmonary patients with simple smartphones for years and demonstrated motion information extracted from phone sensors have predicting power for GOLD levels. In this study, we show that a universal model, trained with about sixty patients 6MWT data, has perfect accuracy at detecting all status from GOLD 0 all the way to GOLD 3. Meanwhile, tree-based models are applied so that we obtained more comprehensive information about the model mechanisms. This gives us a way to better interpret how the predictive model works and how the input features, containing demographics and gait features, interact with each other.

In this study we introduced a potential solution to monitor health status for cardiopulmonary patients, by establishing predictive models with human motion data. We collected 6MWT motion data for sixty-six subjects, including cardiopulmonary patients and healthy subjects. We integrated two tasks: 1) detecting healthy status or unhealthy status and 2) detecting different GOLD stages within unhealthy status; and showed that universal models can predict GOLD status with very high accuracy.

The predictive models are more robust at subject level, after majority voting on sample level predictions. Even the linear SVM model, performing the worst with only 71% accuracy, only have two false predictions at subject level. Moreover, the false predictions are for a

GOLD 1 patient and a GOLD 2 patient. As the most confusing categories for a motion based detector, GOLD 1 and GOLD 2 patients have similar level of mobility, which can be reflected by their 6MWDs. A GOLD 1 patient can perform poor in a 6MWT with bad behavior and a GOLD 2 patient can achieve much higher 6MWD after rehab training.

We showed that RBF-SVM model leads to much better performance than a linear SVM model within the range of GOLD 1 to GOLD 3. In this study, after involving healthy subjects data, RBF-SVM model still obtains close-to-perfect predicting results, even at the sample level.

A major novelty in this study is that we applied the tree-based pattern mining technique, DPClass, to provide more insights about model mechanisms. The decision tree model initially demonstrated that demographic information determines how models will select the gait feature thresholds. Additionally, DPClass is designed to not only build an accurate model but also generate more explainable features using discriminative patterns. The model complexity of DPClass is much smaller than RBF-SVM. For example, DPClass only uses 30 dimensions in its final model, while RBF-SVM will use N dimensions (N is the number of patients). It is not surprising that RBF-SVM has a slightly better accuracy due to its much higher model complexity. However, RBF-SVM is more like a blackbox without explanations. With the explicit patterns, DPClass model shows how demographic information interacts with gait features.

Age is the dominant demographic feature in both decision tree model and DPClass model. Young subjects have less probability to have cardiopulmonary diseases, while also moving more steadily. For cardiopulmonary patients, different age ranges indicates different thresholds on gait features for classification. Besides age, both height and weight plays significant roles in model prediction. This also matches why BMIs are medically important. Height has been adjusted in pulmonary function tests by demographics built into spirometers, but weight has not. So weight should participate more in demographic calibration than height. We observed that both the decision tree model and DPClass model shows weight is more important than height, which matches the medical assumption above. Unlike age, height and weight, sex turns out to be less important, which makes sense because older males and females do not perform differently in steady-pace walking, as reflected on their 6MWD records. The tree structure and discriminative rules provides potential answers to several questions we found in our previous studies, which we will discuss in the next section.

6. Why Model Interpretation

We have been studying using smartphone sensors to monitor human health since 2011. The results were progressively better, but the models always had some errors that we tried to explain with clinical variation. By increasing the number of subjects and refining the techniques, the universal predictive model is now able to detect status from GOLD 0 to GOLD 3 within the cardiopulmonary status. However, several questions related to the mechanisms of model prediction still remain, causing inaccuracies in the models.

We discuss below how the significant features in the discriminative rules explain these questions. As Table 4 shows, there are 5 features (out of 13 total) which occur in at least 10 rules. These are the most significant features, which are automatically identified by the pattern model, the next most significant appears in only 5 rules. These features are discussed, showing that in fact each of the questions raised by our previous studies depend on these significant features.

Age dominates healthy and unhealthy status classification. We showed the initial results for detecting healthy and unhealthy status with six subjects, three healthy and three unhealthy¹⁸. The result shows that healthy and unhealthy subjects are clearly distinguished by their gait speed. However, such correlation is not strong enough to show the causality. Age bias occurs in our dataset and most healthy subjects are younger than patients. We know that young people generally walk faster than old people. Thus, we always question on whether age is the dominant factor for distinguishing the two groups, not health status.

The tree structure in the decision tree model and the discriminative rules in the DPClass model both demonstrate that age matters. The decision tree will directly classify all subjects into healthy category (GOLD 0) if age is younger than 49.5 years. However, this does not mean that the predictive models cannot detect healthy subjects above age 49.5. Actually, we have one healthy subject in the same age range (>60) as cardiopulmonary patients, and all his ten-second samples are correctly classified into GOLD 0. DPClass also gives more detailed age thresholds for each discriminative rule. These rules are not only for detecting healthy versus unhealthy status but different severity levels. Thus, age is not a factor that causes bias in our dataset but a demographic factor that essentially affects model prediction.

BMI significantly distinguishes demographic cohorts

We had published a pilot study that discusses demographic effects on model prediction for different GOLD stages¹⁶. With an empirically picked demographic cohort of sex (female) and age (50-64), the predicting accuracy significantly increases. However, this is not the full story for demographic effects. BMI, which is a combined effect of height and weight, is shown to be significant for model predicting, after we merged the demographic information and gait features into the model input, leading to a close-to-perfect prediction. However, then we did not know how BMI affects model prediction. In the current study, we observed that weight dominates decision making for several nodes in the decision tree. In fact, height (12/30) and weight (18/30) both appear frequently in the discriminative rules for the DPClass model. All these evidences show that BMI plays significant role in GOLD prediction.

However sex does not make significant difference for model prediction from the result of decision tree and DPClass. Sex does not dominate any tree nodes and only contributes in two out of thirty discriminative rules. There are two potential reasons: 1) Sex difference does not affect 6MWDs significantly ($p > 0.05$). Unlike fatigue tests which test the upper bound of mobility, natural walking of males and females are less distinguishable because the gait indicators do not vary much. 2) Pulmonary function test has already adjusted the predicted FEV1% values by demographics including age, sex, height and race. Unlike age and height, sex or gender is a simple binary indicator, so easy to adjust the predicted FEV1% against sex

across the whole population. Also, since weight is not in the adjustment of PFTs but height is, raw weight value has stronger predictive power, so weight plays a more important role in our model.

MeanAcc more important than Cadence

We carefully studied on natural walking and showed that simple step counting or walk distance is not good enough for cardiopulmonary status prediction. Neither cadence nor 6MWD can be directly used to distinguish GOLD stages¹⁰. Since cadence, defined the number of strides in a minute, is an essential parameter in gait analysis for general purpose, we still assume that cadence is a distinguishing factor⁸. However, both the decision tree structure and discriminative rules show that cadence is not comparatively significant as the mean acceleration, which is a feature continuously measuring 3D spatial motion of human body. We observed that cadence does not dominate tree nodes but mean acceleration (MeanAcc) does, and cadence participates in only two out of thirty discriminative rules while mean acceleration participates in fifteen out of thirty discriminative rules. Thus, a predictive model for cardiopulmonary status requires more information about motion than simple steps or speed. Note this implies that fitness devices, which only compute steps, cannot predict function, unlike smart phones with appropriate software, which can detect the slowdown and speedup patterns during natural walking.

RMS is more important than StdAcc

Future prediction is critical to prescribing treatments. This involves predicting “risk”, which is the change in status. We have made efforts for risk assessment, which is one step beyond status detection. In particular, we developed a model for predicting transitions in-between two consecutive GOLD stages¹³. In general, instability reflects risks. It means that if the status is more unstable, the higher the risk that this patient will have status change in the future. Both the standard deviation of acceleration (StdAcc) and root mean square of acceleration (RMS) reflect the stability of walking. RMS measures the overall acceleration while StdAcc measures the offside around mean acceleration. Nishiguchi *et al.* claimed that RMS is a better measure for stability for human walking¹⁵. Here in this study we attested that RMS is more important than StdAcc for model prediction. RMS dominates two tree nodes in decision tree model while StdAcc does not dominate any tree nodes. Meanwhile, we also observed that RMS contributes to eleven of thirty discriminative rules and StdAcc only contributes to five of them. Thus, for a potential risk assessment for clinical instability, the root mean square of acceleration could be a better indicator than the standard deviation of acceleration.

7. Future Work

Passively monitoring patient status will be the promising application of this study, requiring additional techniques to take the motion collection out of clinic into the home. Besides a precise prediction for health status, *e.g.* GOLD stages, passively collecting sensor data from phones and correctly retrieving good walking motion are also important components. The next stage is already implemented with IRB approval at NorthShore HealthSystem, but just beginning to be tested as this paper is written. On the next stage, we will recruit

cardiopulmonary patients for the passive monitoring experiments. These will contain two stages, a clinical examination in rehabilitation and a take-home field trial during daily activities. The clinical examination sets up a baseline for this patient, containing all required tests to obtain motion and status information. The take-home field trial is a follow-up step after the patient completes clinical examination. We will give this patient a smartphone, with a passive version of MoveSense installed, that continuously monitors daily motion at home and transmits sensor data to our server. Our passive monitoring software optimizes the collection task and battery usage, by only recording when the phone is moving and only storing motion that represents good walking. The patients will simply carry their phones during normal activities, while producing multiple walking sessions as inputs to the predictive models. With such good walking data, we will be able to obtain a daily status and risk change for all patients. This new clinical experiment will evaluate whether we can model pulmonary function with as high accuracy in the home as we have already done in the clinic. After this next step, we are planning to begin a full-scale trial with thousands of patients, to support population measurement of health status from carried phones with clinical accuracy.

8. Conclusion

In this study, we focus on the validity of predictive models for monitoring health status, including both model performance and model interpretation. We carefully prepare motion dataset covering status from GOLD 0 (healthy), GOLD 1 (mild), GOLD 2 (moderate), all the way to GOLD 3 (severe). We then train models with different supervised learning techniques to demonstrate that human motion information has potential capacity to predict pulmonary functions for cardiopulmonary diseases. A universal model, which accepts demographic features and gait features, is able to obtain perfect prediction result. Also, the tree-based models: decision tree and DPClass, provide significant insights into how demographics interacts in model prediction. This gives us a potential solution to automatically detect different demographic cohorts within car-diopulmonary patients.

Acknowledgments

At the University of Illinois at Urbana-Champaign, the Institute for Genomic Biology provided facilities for software development and data analysis. The research project was sponsored in part by grant 1U54GM114838 awarded by NIGMS through funds provided by the trans-NIH Big Data to Knowledge (BD2K) initiative (www.bd2k.nih.gov) and the National Science Foundation IIS-1320617, and IIS-1354329. At Carle Foundation Hospital in Urbana Illinois, the Biomedical Research Institute helped test the patients. Dustie Mitchell, RN, provided medical supervision during patient testing. Debbie Brusveen, RN, provided access to the Pulmonary Rehabilitation facility, where the tests were run. Sujay Bangarulingam, MD, was the pulmonary physician responsible for IRB 497221. At NorthShore University HealthSystem, the Department of Respiratory Therapy at Evanston Hospital carried out patient testing and phone recording, with Deanna Close, RN, and Nicholas Fulara supervising the patient rehabilitation. These clinical experiments were protected by NorthShore IRB EH15-025, with Shashi Bellam, MD, as the responsible pulmonary physician. The Center for Biomedical Research Informatics supported the clinical experiments, thanks to their Vice President, Jonathan Silverstein, MD. We acknowledge Brian Edwards and Catherine Zhu for handling phone data and medical records inside NorthShore University HealthSystem.

References

1. Schatz, BR., Berlin, RB. Healthcare Infrastructure: Health Systems for Individuals and Populations. Springer London Limited; 2011. Series in Health Informatics

2. Smith A. U.S. Smartphone use in 2015. Pew Research Internet Project. 2015
3. Broaddus, VC., Mason, RC., Ernst, JD., King, TE., Jr, Lazarus, SC., Murray, JF., Nadel, JA., Slutsky, A., Gotway, M. Murray & Nadel's Textbook of respiratory medicine. Elsevier Health Sciences; 2015.
4. Annegarn J, Spruit MA, Savelberg HH, Willems PJ, Boud Cvd, Schols AM, Wouters EF, Meijer K. Differences in walking pattern during 6-min walk test between patients with copd and healthy subjects. PLOS ONE. 2012; 7(5):e37329. [PubMed: 22624017]
5. Jehn M, Schmidt-Trucksäess A, Schuster T, Hanssen H, Weis M, Halle M, Koehler F. Accelerometer-based quantification of 6-minute walk test performance in patients with chronic heart failure: applicability in telemedicine. Journal of Cardiac Failure. 2009; 15(4):334–340. [PubMed: 19398082]
6. ATS. ATS statement: Guidelines for the six-minute walk test. American Journal of Respiratory and Critical Care medicine. 2002; 166(9):111–117. [PubMed: 12091180]
7. Wiener, C., Fauci, A., Braunwald, E., Kasper, D., Hauser, S., Longo, D., Jameson, J., Loscalzo, J., Brown, C. Harrison's Principles of Internal Medicine Self-Assessment and Board Review 18th Edition. McGraw Hill Professional; 2012.
8. Whittle, MW. Gait analysis: an introduction. Butterworth-Heinemann; 2014.
9. Shang, J., Tong, W., Peng, J., Han, J. Dpclass: An effective but concise discriminative patterns-based classification framework; Proceedings of the SIAM International Conference on Data Mining; 2016.
10. Juen J, Cheng Q, Schatz B. A natural walking monitor for pulmonary patients using mobile phones. Biomedical and Health Informatics, IEEE Journal. 2015; 19(4):1399–1405.
11. Cheng Q, Juen J, Bellam S, Close D, Fulara N, Silverstein J, Schatz B. Predictive models for pulmonary function using motion analysis from phone sensors. AMIA Annual Symposium Proceedings. 2016 submitted.
12. Cheng, Q., Juen, J., Schatz, BR. Proceedings of the 5th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics. ACM; 2014. Using mobile phones to simulate pulse oximeters: gait analysis predicts oxygen saturation; p. 331-340.
13. Cheng Q, Juen J, Hsu-Lumetta J, Schatz B. Predicting transitions in oxygen saturation using phone sensors. Telemedicine and e-Health. 2015
14. Lu, H., Yang, J., Liu, Z., Lane, ND., Choudhury, T., Campbell, AT. Embedded Networked Sensor Systems. ACM; 2010. The Jigsaw continuous sensing engine for mobile phone applications; p. 71-84.
15. Nishiguchi S, Yamada M, Nagai K, Mori S, Kajiura Y, Sonoda T, Yoshimura K, Yoshitomi H, Ito H, Okamoto K, et al. Reliability and validity of gait analysis by android-based smartphone. Telemedicine and e-Health. 2012; 18(4):292–296. [PubMed: 22400972]
16. 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
17. Cortes C, Vapnik V. Support-vector networks. Machine learning. 1995; 20(3):273–297.
18. Cheng, Q., Juen, J., Li, Y., Prieto-Centurion, V., Krishnan, JA., Schatz, BR. Proceedings of the 4th International Conference on Bioinformatics, Computational Biology and Biomedical Informatics. ACM; 2013. GaitTrack: Health monitoring of body motion from spatio-temporal parameters of simple smart phones; p. 897
19. Therneau, TM., Atkinson, B., Ripley, B. rpart: Recursive partitioning. r package version 3.1-46. Computer software program. 2010. retrieved from <http://CRAN.R-project.org/package=rpart>
20. Chen, C., Liaw, A., Breiman, L. Using random forest to learn imbalanced data. University of California; Berkeley: 2004.
21. Tibshirani R. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society Series B (Methodological). 1996:267–288.
22. Friedman J, Hastie T, Tibshirani R. glmnet 1.1: lasso and elastic-net regularized generalized linear models. r package. 2009

Figure 1. Overview of DPClass

Figure 2.

Brief Decision Tree Structure. Each Node represents a specific feature dominating decision making on this step. Decision trees making binary decision trees on each node.

Demographic information of each group by GOLD levels, Patient [Female] shows the number of patient [the number of female patient] in each group. Age, height and weight are in median [minimum - maximum] format.

Table 1

	GOLD 0	GOLD 1	GOLD 2	GOLD 3	Overall
Patients [Female]	9[6]	23[14]	18[8]	9[3]	59[31]
Age (year)	28[20-60]	65[43-81]	76[61-95]	75[55-85]	68[20-95]
Height (m)	1.72[1.57-1.88]	1.68[1.47-1.88]	1.68[1.24-1.83]	1.68[1.55-1.83]	1.68[1.24-1.88]
Weight (kg)	75.8[51.3-104.3]	85.28[49.4-126.1]	78.93[48.08-117.03]	81.65[45.36-118.39]	80.51[45.36-126.1]

Table 2

Model Validation. We validate four different predictive models by 10-fold cross validation. Each column represents the prediction accuracy for samples in the corresponding category. The accuracy is calculated by the portion of number of correctly predicted samples in the total number of samples in this category.

	GOLD 0	GOLD 1	GOLD 2	GOLD 3	Overall
Linear SVM	0.9959	0.8069	0.5432	0.6064	0.7140
RBF-SVM	0.9938	0.9965	0.9874	0.9888	0.9920
Decision Tree	0.9959	0.9730	0.9329	0.8656	0.9456
DPClass	1.0000	0.9689	0.9262	0.9712	0.9597

Table 3

Confusion Matrices of Model Validation Results. In this table we show the results of model prediction. GOLD n represents the actual GOLD stage and Pred GOLD n represents the predicted GOLD stage. Numbers of samples on diagonals represents numbers of correctly predicted samples in each category. Note these are only 10 second window samples, but only a very few are incorrect predictions.

	Linear SVM				RBF-SVM			
	GOLD 0	GOLD 1	GOLD 2	GOLD 3	GOLD 0	GOLD 1	GOLD 2	GOLD 3
Pred GOLD 0	483	4	0	0	482	0	0	0
Pred GOLD 1	2	1166	352	79	3	1440	10	0
Pred GOLD 2	0	243	648	167	0	5	1178	7
Pred GOLD 3	0	32	193	379	0	0	5	618
	Decision Tree				DPClass			
	GOLD 0	GOLD 1	GOLD 2	GOLD 3	GOLD 0	GOLD 1	GOLD 2	GOLD 3
Pred GOLD 0	484	21	0	0	485	7	0	0
Pred GOLD 1	0	1406	68	10	0	1400	83	6
Pred GOLD 2	1	16	1113	74	0	35	1105	12
Pred GOLD 3	0	2	12	541	0	3	5	607

Table 4

Proportion of Rules Containing Each Feature. We count the number of appearance of a specific feature in all top 30 rules.

Feature	Portion of Rules
Age	24/30
Weight	18/30
Height	12/30
Sex	2/30
Cadence	2/30
MeanAcc	15/30
StdAcc	5/30
AC	3/30
CV	3/30
MCR	1/30
RMS	11/30
PF	4/30

Table 5

Discriminative Patterns in DPClass Model. A discriminative pattern is a conjunctive clause containing several conditions, for example “(Age < 80.5) AND (Height > 1.5)”. Each discriminative pattern represents a rule for making binary decision.

Rank	Discriminative rules
1	(Age < 78.5) AND (Weight >= 63.28) AND (MeanAcc >= 9.87816) AND (MeanAcc < 10.1604)
2	(Age < 81.5) AND (Height >= 1.537) AND (Weight >= 59.875) AND (MeanAcc >= 9.76779) AND (MeanAcc >= 9.88939) AND (StdAcc < 2.79557)
3	(Age < 49.5) AND (MeanAcc < 9.86086)
4	(Weight >= 63.28) AND (MeanAcc >= 9.74457) AND (MeanAcc >= 9.86863) AND (RMS < 1.04414) AND (RMS >= 1.0138)
5	(Age < 80.5) AND (Weight >= 63.28) AND (MeanAcc >= 9.67007) AND (StdAcc >= 1.09936) AND (RMS < 1.05517)
6	(Weight >= 67.815) AND (RMS >= 1.01061)
7	(Age >= 51) AND (Age >= 66.5) AND (Weight < 117.71) AND (PF < 2.74542)
8	(Age >= 49.5) AND (Weight >= 61.01) AND (RMS >= 0.993072) AND (RMS >= 1.01945) AND (RMS < 1.05535)
9	(Age < 63.5) AND (Height >= 1.635) AND (Weight < 115.215) AND (MeanAcc < 9.76779)
10	(MeanAcc >= 9.78772) AND (RMS < 1.06968)
11	(Age >= 66.5) AND (Height < 1.7763) AND (AC >= 0.507291)
12	(Age >= 66.5) AND (Height < 1.775) AND (Weight >= 56.475) AND (MeanAcc < 9.88532)
13	(Age >= 66.5) AND (Age < 69.5) AND (Height < 1.7763) AND (AC >= 0.507291)
14	(Age >= 52.5) AND (Height < 1.7513) AND (cadence >= 0.675) AND (StdAcc >= 0.745093) AND (StdAcc < 1.2303) AND (RMS < 1.01391)
15	(Age < 49.5)
16	(Age >= 66.5) AND (Height >= 1.7763) AND (Weight < 89.815)
17	(Age >= 77.5) AND (Weight < 56.475) AND (CV >= 0.0941185)
18	(Age >= 60.5) AND (Weight >= 55.565) AND (Weight >= 78.925) AND (MeanAcc < 9.76984) AND (StdAcc < 1.32393) AND (PF < 1.84692)
19	(Height < 1.74) AND (Height >= 1.395) AND (cadence >= 0.725) AND (RMS < 1.0102) AND (CV < 0.114057) AND (CV >= 0.0844477)
20	(Height < 1.665) AND (Weight < 58.97) AND (AC >= 0.681601)
21	(Age < 64) AND (Weight >= 48.31) AND (MeanAcc < 9.71001)
22	(Age >= 49.5)
23	(Age >= 66.5) AND (Height >= 1.6637) AND (Weight >= 56.475) AND (Weight < 93.215) AND (MeanAcc >= 9.88532)
24	(Age >= 33) AND (Age < 66.5) AND (MeanAcc >= 9.7852) AND (StdAcc >= 1.11517) AND (StdAcc < 2.99304)
25	(Age >= 52.5) AND (Height >= 1.6505) AND (MeanAcc >= 9.72688) AND (MeanAcc < 9.88382)
26	(Age < 66.5) AND (Height >= 1.653) AND (Weight >= 56.475) AND (Weight < 108.865) AND (MeanAcc < 9.88532) AND (RMS < 1.00835)
27	(Weight >= 65.545) AND (Weight < 122.015) AND (Weight >= 94.57) AND (MeanAcc >= 9.98547) AND (RMS >= 1.01831) AND (CV >= 0.187085)
28	(Age < 56) AND (Weight < 82.775)
29	(Age >= 49.5) AND (Sex >= 1.5) AND (MCR >= 0.388519) AND (RMS >= 1.00043) AND (PF >= 1.64725)
30	(Age >= 49.5) AND (Sex >= 1.5) AND (RMS >= 1.00043) AND (PF >= 1.64725)