35th Annual International Conference of the IEEE EMBS
Osaka, Japan, 3 - 7 July, 2013

Prediction of Chronic Obstructive Pulmonary Disease Exacerbation using Physiological Time Series Patterns

Yang Xie[1], *Student Member, IEEE*, Stephen J. Redmond[1], *Member, IEEE*, Mas S. Mohktar[2], Tal Shany[1], Jim Basilakis[3], Michael Hession[4], Nigel H. Lovell[1], *Fellow, IEEE*

***Abstract*— Chronic obstructive pulmonary disease (COPD) is responsible for significant morbidity and mortality worldwide. Recent clinical research has indicated a strong association between physiological homeostasis and the onset of COPD exacerbation. Thus the analysis of these variables may yield a means of predicting a COPD exacerbation in the near future. However, the accuracy of existing prediction methods based on statistical analysis of periodic snapshots of physiological variables is still far from satisfactory, due to lack of integration of long-term and interactive effects of the physiological variables. Therefore, developing a relatively accurate method for predicting COPD exacerbation is an outstanding challenge. In this paper, a regression-based machine learning technique was developed, using trend pattern variables extracted from COPD patients' longitudinal physiological records, to classify subjects into "low-risk" and "high-risk" categories, indicating their risk of suffering a COPD exacerbation event. Experimental results from cross validation assessment of the classifier model show an average accuracy of 79.27% using this method.**

I. INTRODUCTION

Chronic obstructive pulmonary disease (COPD) is responsible for significant morbidity and mortality worldwide [1]. The major contributor to COPD morbidity is exacerbation, a worsening of COPD symptoms, and its association with other co-morbidities. Both factors have increased the cost of managing COPD [2]. Therefore, for better management and more timely treatment of COPD patients, an earlier identification of exacerbation risk is needed.

Recent research reveals a strong association between changes of vital physiological parameters and the onset of COPD exacerbation. Heart rate (HR) and peripheral arterial oxygen saturation (SpO2) have been shown capable distinguishing the onset of COPD exacerbation from normal variations in symptoms. As a criterion for screening insufficient blood oxygen saturation [3], an SpO2 value less than 88% is considered a sign of instability in COPD patients [4]. HR is reported to increase as COPD exacerbations worsen. A HR value over 110 BPM (beats per minute) or 20% above baseline is one of the clinical guideline indicators to assess exacerbation severity [1].

In addition, the correlation between low blood pressure and COPD exacerbation-related in-hospital mortality has been reported by Edwards et al. [5]. Body temperature is also a critical sign of patient' wellbeing. One study reported that over one-third of exacerbations requiring hospitalisation had fever as an indicator [6]. COPD patients often loose weight as a result of exacerbation of their disease [7].

Most of the above work, if not all, is based on statistical analysis of periodic snapshots of physiological parameters weekly or monthly. Whether patterns from a continuous physiological data stream, such as long-term data trends may have a predictive role, is rarely studied. Therefore, a lack of investigation into the predictive power of longitudinal physiological data trends has motivated the research contained herein, to better predict the onset of COPD exacerbation.

In this paper, a trend detection technique, previously developed by our research group [8] was employed to automatically apply piecewise linear fits to COPD patients' longitudinal physiological measurement records. A number of features were later extracted from these trend fits and used to train a classifier model to predict the onset of exacerbation.

II. METHODS

A. Data set

The data was obtained from seven COPD patients, aged between 63-87 years old, who lives in the western suburbs of the city of Sydney in Australia. Data was collected by staff at Blacktown Hospital, Sydney, from April 2009 to October 2010, using a TeleMedCare Health Monitor (TeleMedCare Pty. Ltd., Sydney, Australia). For each subject, six physiological parameters were measured daily for approximately one year: weight, diastolic blood pressure (DBP), systolic blood pressure (SBP), HR, SpO2 and temperature. The measurements of each type of physiological parameter form a time series, in which trends are later detected. In addition, standardised health questionnaires were used to assess the symptoms of illness and the mood of patients. Treatment with medication was also monitored. It should be pointed out that, in this study, the spirometry measurements were not consistently recorded since the patients were not encouraged to perform spirometry in an unsupervised environment by the staff of the hospital. This results in a limitation of the work presented here.

* This work was supported by Austrilian Research Council Linkage Grant.

[1] Graduate School of Biomedical Engineering, University of New South Wales, Sydney, New South Wales, 2052, Australia. n.lovell@unsw.edu.au

[2] Department of Biomedical Engineering, Faculty of Engineering, University of Malaya, 50603 Kuala Lumpur, Malaysia.

[3] School of Computing and Mathematics, University of Western Sydney, Campbelltown, Sydney, New South Wales, Australia.

[4] Blacktown and Mt Druitt Hospital, Sydney, New South Wales, Australia.

The authors would like to acknowledge the valuable contributions made by David Pryce and Rowena Galang during the data collection and analysis phases of the trial.

Before detecting trends in the physiological parameter time series, the raw physiological measurement data was pre-processed to remove possible outliers. For any repeated measurements on the same day, an average for that day was used.

B. Trend detection

To identify trends in the physiological time series data for each subject, over approximately one whole year, a piecewise regression algorithm, introduced in [8], was employed. The piecewise linear trend detection algorithm has been shown to perform well in detecting underlying trends in data corrupted by noise. The location of the K breakpoints in the piecewise linear fit are selected using a backward selection search, sequentially removing the breakpoint which gives the least increase in the mean squared error of the fit, but forbidding the removal of any breakpoint which would result in any single error exceeding a predetermined limit, termed e_{max}, which is heuristically chosen as the fraction of the standard deviation of the data over the entire year. Fig. 1 demonstrates a sample of weight measurements across approximately one year.

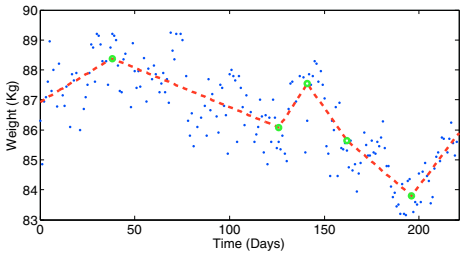

Fig. 1. A sample of approximately one year of weight measurements. The blue dots are the weight data points. The green circles are the detected breakpoints. The red dash line is the piecewise trend detected.

C. Feature extraction

To estimate a daily health status, an "online" scheme was used. When a new day's physiological measurement data arrives, the trend fit is updated using the above technique and the trend pattern features are extracted from the trend for the day. The measurement data is recorded in the format of (t, y), in which t is the abscissa (date) and y is the ordinate (measurement value). Accordingly, nine basic features were generated for each day:

1 Measurement value: the value of the physiological measurement on the target day (y_{target}), that is the most recent day; if missing, an interpolated value is used instead.

2 Time to last breakpoint: the time difference between the target day (t_{target}) and last breakpoint day (t_{lastBP}).

$$t_{diff} = t_{target} - t_{lastBP} \tag{1}$$

3 Slope of trend segment where the target day is located.

$$slope = \frac{y_{target} - y_{lastBP}}{t_{target} - t_{lastBP}} \tag{2}$$

4 Absolute change over segment.

$$|y_{diff}| = |y_{target} - y_{lastBP}| \tag{3}$$

5 Standard deviation of measurements (y_i) over the time interval since the last breakpoint. μ is the mean value of y_i.

$$\sigma_{segment} = \sqrt{E\{(y_i - \mu)^2\}} \tag{4}$$

6 Standard deviation of segment after detrending: standard deviation of the difference between the values of real measurement (y_i) and the values of trend estimate ($\hat{y}_i$):

$$\sigma_{detrend} = \sqrt{E\{(y_i - \hat{y}_i)^2\}} \tag{5}$$

7 σ_S: a statistical parameter returned by the trend detection technique, which is estimated by finding the standard deviation of the data within a short-term time window.

8 σ_L: another statistical parameter returned by the trend detection technique, which is calculated as the standard deviation of the entire signal.

9 σ_S/σ_L: the ratio of σ_S and σ_L. An indicator of signal-to-noise ratio.

Thus 54 features were extracted (9 basic features $\times$ 6 physiological parameters). Furthermore, the squared and cubed values were also added into the feature pool to expand the 54 features to a total of 162 features. These transformed features were included to ensure quadratic and/or cubic non-linear relationships between the features and the exacerbation risk could be captured by the model.

D. Reference standard

In order to build a predictive model, to identify a possible exacerbation event, a reference standard indicating each subject's health status on certain days was constructed using the information in the symptoms and medications questionnaires (e.g., increase in sputum amount/medication dosage)[2]. In this reference, for the days when the questionnaire data and enough physiological data were available, one of two categorical labels was assigned to the patient's health condition: "low-risk" or "high-risk". This reference standard was later used in the classifier training phase, to fine-tune the performance of the classifier, before independent validation was performed. Since the objective is to estimate the daily health status of each subject, the features and references were aggregated at the level of a single day to generate modelling data sets. This resulted in 342 days worth of data from all seven subjects.

E. Logistic regression classifier

Logistic regression is based on a simple logistic function defined by the formula:

$$P(t) = \frac{1}{1 + e^{-t}} \tag{6}$$

This classifier performs very well for predicting the outcome of a binary variable dependent on one or more predictor variables. That is, t is assumed to be a weighted sum of the dependent variables (in this case the extracted features from

Section II-C). Training the classifier involves determining the weights of this sum to generate t, such that $P(t)$ best matches the training data labels. Considering in our case, the outcomes would finally fall into two categories: "low-risk" or "high-risk", a logistic classifier is a reasonable choice.

F. Feature subset selection

Subset selection was employed to find the near optimal subset of features to achieve the highest prediction accuracy. This strategy resulted in an elimination of features from the progressively developed model. In this paper, a forward-backward floating search strategy was used.

The forward search attempts to find the optimal subset of features from the pool of available candidate features. It starts with single feature from all available features, and then sequentially adds into the model the feature that most improves the prediction accuracy. After a feature is selected, removal of a feature from the current set of selected features is attempted. The process of possible feature addition, followed by possible feature removal is iterated until the selected feature set converges.

G. K-fold cross validation

In this study, a double-loop ten-fold cross-validation scheme, with inner and outer nested loops, was used.

The outer loop is used to assess the generalised performance of the classifier model. In the outer loop, data was repeatedly and randomly divided into two sets: modelling data set (nine folds using 90% of the 342 days = 308 days) and validation data set (one fold using 10% of the 342 days = 34 days). The 34 days of validation data were withheld and did not appear in the model selection process, while the 308 days of training data were used in the inner loop for model selection.

An inner loop was used to perform feature selection. In the inner loop, the modelling data was again repeatedly and randomly divided into two subsets: test data (one fold using 10% of the 308 days = 31 days) and training data (nine folds using 90% of the 308 days = 277 days). Again for the inner loop, the test set was withheld from the training set to later test with for each of the 10 cross validation folds. The classifier was trained using the remaining 277 days of training data. Afterwards, the withheld fold of 31 days of test data was then reintroduced for classification. This was repeated ten times until each fold had been test data. The average result across all 10 folds test data was used as the feature selection criteria.

In summary, the inner loop is used to help select the optimal feature subset. The outer loop training data is used for training a "best" model based on those selected features. The outer loop validation data is later applied to the "best" model to test it, the result of which indicates how the "best" model performs on unseen data. To minimize the effect of sampling bias when assigning data to folds, both the inner and outer ten-fold cross-validations were repeated ten times. Data was randomly reassigned to folds on each repeat.

III. RESULTS

Table I lists the performance metrics for the proposed method. Shown are the aggregated results of the outer most 10-fold cross validation loop (each row in the table represents one of the 10 repetitions). The result of each repetition (each row) is calculated by averaging the results of each of the 10 folds for that repeat.

The logistic regression classifier initially returned the estimated health risk score as a probability, which was later rounded to 0 or 1. The confusion matrix (CM), accuracy (Acc.) and Cohen's kappa (κ) metrics were calculated comparing the binary estimated health risk score against the reference health risk score. The area under receiver operating characteristic (ROC) curve was also calculated and presented (abbreviated AUC).

Fig. 2 demonstrates the box plot of the estimated health risk score against the reference health risk score for a sample repetition. On each box, the central mark is the median; the edges of the box are the 25^{th} and 75^{th} percentiles.

The features selected in each "best" model were later pooled together, in which 100 sets (10 repeats $\times$ 10 folds of the outer validation loop) of "optimal" features were aggregated. Fig. 3 shows a pie chart of frequency of appearance of these features categorized by trend pattern feature type. Fig. 4 demonstrates a similar pie chart, but categorized by physiological parameter type.

TABLE I

A SUMMARY OF THE STATISTICAL RESULTS OF THE TEN OUTER LOOP REPETITIONS OF LOGISTIC CLASSIFIER.

No	AUC	κ	Acc. (%)	CM	
1	0.8215	0.5355	76.90	119 39	40 144
2	0.8579	0.5841	79.24	127 31	40 144
3	0.8227	0.5731	78.65	128 30	43 141
4	0.8736	0.6592	83.04	130 28	30 154
5	0.8424	0.5659	78.36	124 34	40 144
6	0.8446	0.6061	80.41	125 33	34 150
7	0.8613	0.6246	81.29	129 29	35 149
8	0.8424	0.5416	77.19	120 38	40 144
9	0.8532	0.5965	79.82	130 28	41 143
10	0.8525	0.5538	77.78	122 36	40 144
Average	0.8472	0.5840	79.27	125.4 32.6	38.3 145.7

IV. DISCUSSION AND CONCLUSION

A method for predicting daily COPD exacerbation risk has been developed using the trend pattern features obtained from longitudinal physiological measurement data. The method has been evaluated using cross validation. Table I shows that

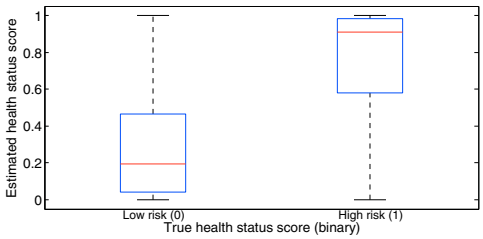

Fig. 2. Box plot of the estimated score risk score against the reference health risk score (logistic regression probability) for a sample repetition (the 10^{th} repetition in Table I).

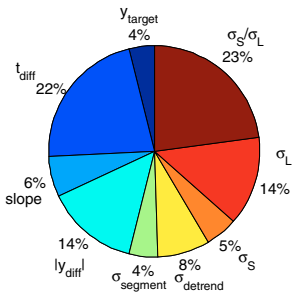

Fig. 3. Pooled pie chart of selected features categorized by trend pattern feature type. In total, 100 sets of "optimal" features were aggregated.

the model has a fair average accuracy (79.27%) in classifying between low-risk and high-risk days. The results of all the ten repetitions show that the model selection method in the inside loops is relatively stable with accuracies ranging from 76.9% to 83.04%. The stability can also be viewed from other statistical indicators, such as AUC, κ and CMs. From Fig. 2, a box plot of a sample repetition is displayed. In the low-risk box, the median value is approximately 0.2, while the high-risk box median value is approximately 0.9. This further supports a good separation of risk categories using this classification scheme.

Pooled pie charts of the selected features from all "best" models are presented in Fig. 3 and Fig. 4. The pie chart in Fig. 3 is categorized by the type of the trend pattern feature, in which we can see that the most-frequently-appear trend pattern features are t_{diff} and σ_S/σ_L. The piecewise linear trend detection technique aims to capture incipient changes in the longitudinal data. The t_{diff} feature is the days since last break point day which may imply an impending

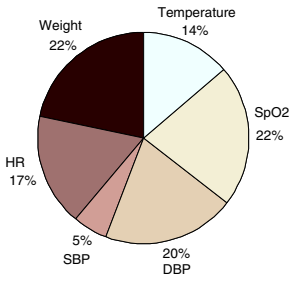

Fig. 4. Pooled pie chart of selected features categorized by physiological parameter type. In total, 100 sets of "optimal" features were aggregated.

health status change. The σ_S/σ_L feature could be considered as an indicator of signal-to-noise ratio. For poor signal-to-noise, σ_S/σ_L approaches 1. For a large signal-to-noise ratio, σ_S/σ_L approaches 0. The implication might be large variations in parameters imply instability. In Fig. 4 the pie chart is divided by the type of physiological measurement, which shows a relatively even division between the five of the six physiological measurements, with systolic blood pressure chosen less often. Both weight and SpO2 appeared most frequently, each accounting for 22% of the selected features, but not much different from HR (17%), DBP (20%) and temperature (14%). However, it is interesting to note that SBP is only selected 5% of the time. Edwards et al. reported that a low blood pressure was associated with COPD exacerbation [5], which could be identified as a SBP less than 90mmHg or a DBP less than 60mmHg. Although, both DBP and SBP are important indicators of low blood pressure, it seems that DBP provides more predictive power than SBP.

Although the six available physiological measurements gave a good predictive performance, some limitation still lies in the lack of other important physiological measurements of the forced spirometry manoeuvre, such as forced expiratory volume in one second (FEV1), which has been previously reported to be associated with the severity of COPD [4]. Future work will focus on applying the method developed in this paper to other telehealth databases with more complete longitudinal physiological measurement records, in particular those which contain forced and relaxed spirometry measures, which are expected to further boost the prediction performance demonstrated here.

REFERENCES

[1] F. Gómez and R. Rodriguez-Roisin, "Global initiative for chronic obstructive lung disease (GOLD) guidelines for chronic obstructive pulmonary disease," *Current opinion in pulmonary medicine*, vol. 8, no. 2, pp. 81–86, 2002.

[2] M. S. Mohktar, "A decision support system for the home management of patients with chronic obstructive pulmonary disease (COPD) using telehealth," Ph.D. dissertation, Graduate School of Biomedical Engineering, Faculty of Engineering, University of New South Wales, 2012.

[3] M. Gryay, E. Ceylan, T. Gnay, S. Karaduman, F. Bengi, I. Parlak, M. Ciek, and A. Cimrin, "Can spirometry, pulse oximetry and dyspnea scoring reflect respiratory failure in patients with chronic obstructive pulmonary disease exacerbation?" *Medical Principles and Practice*, vol. 16, pp. 378–383, 2007.

[4] D. McKenzie, M. Abramson, A. Crocket, N. Glasgow, S. Jenkins, C. McDonald *et al.*, "The COPD-X plan: Australian and New Zealand guidelines for the management of chronic obstructive pulmonary disease," 2008.

[5] L. Edwards, K. Perrin, M. Wijesinghe, M. Weatherall, R. Beasley, and J. Travers, "The value of the CRB65 score to predict mortality in exacerbations of COPD requiring hospital admission," *Respirology*, vol. 16, pp. 625–9, 2011.

[6] D. Lieberman, O. Shmarkov, Y. Gelfer, R. Varshavsky, and D. V. Lieberman, "Prevalence and clinical significance of fever in acute exacerbations of chronic obstructive pulmonary disease," *Eur J Clin Microbiol Infect Dis*, vol. 22, pp. 75–8, 2003.

[7] M. Barnett, *Chronic Obstructive Pulmonary Disease in Primary Care*. West Sussex, England, John Wiley and Sons, Ltd, 2006.

[8] S. Redmond, J. Basilakis, Y. Xie, B. Celler, and N. Lovell, "Piecewise-linear trend detection in longitudinal physiological measurements," in *Proc. 31th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 3413 –3416, 2009.