2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Remote COPD Severity and Exacerbation Detection Using Heart Rate and Activity Data Measured from a Wearable Device

Abhishek Tiwari[1], Salaar Liaqat[2], Daniyal Liaqat[2], Moshe Gabel[2], Eyal de Lara[2], and Tiago H. Falk[1]
[1] Institut national de la recherche scientifique, INRS-EMT, Montreal, Canada
[2] University of Toronto, Canada

***Abstract*— Chronic obstructive pulmonary disease (COPD) is one of the leading causes of human mortality worldwide. Traditionally, estimating COPD severity has been done in controlled clinical conditions using cough sounds, respiration, and heart rate variability, with the latter reporting insights on the autonomic dysfunction caused by the disease. Advancements in remote monitoring and wearable device technologies, in turn, have allowed for remote COPD monitoring in daily life conditions. In this study, we explore the potential for predicting COPD severity and exacerbation using a low-cost wearable device that measures heart rate and activity data. We collected smartwatch sensor data from 35 COPD patients over a period of three months. Our evaluation shows that future trajectory of the disease can be predicted using only the first few days of continuous unobtrusive wearable data collected from COPD patients. Using features extracted from wearable device an Isolation Forest was able to predict exacerbation with an area under curve (AUC) 0.69 thus showing improvement over a random choice classifier.**

I. Introduction

Chronic obstructive pulmonary disease (COPD) is one of the leading causes of death worldwide [1]. The disease is characterized by a progressive blockage of airways leading to shortness of breath and decreased quality-of-life for affected individuals [2]. Exacerbation of these chronic conditions often leads to frequent hospitalizations and increased healthcare costs. Recent focus towards remote monitoring of the disease has shown reduced costs with improvements in patient quality-of-life [3]. Additionally, focus towards remote monitoring of chronic disease patients has received a renewed push due to the recent COVID-19 pandemic, which has greatly burdened the healthcare systems worldwide [4]. Remote health monitoring is further enabled by recent advances consumer sensing technologies for wearable devices allowing for long-term, unobtrusive, and continuous acquisition of biomedical data [5].

As a result, several studies have made use of wearable devices to monitor COPD. These include smartphones for monitoring coughing, lung sounds, respiratory rate and also for gait and activity sensing [6]. Heart rate (HR) and Heart Rate Variability (HRV) measured using smartwatches and/or chest-straps [6]–[8] have also been a popular tool for COPD monitoring [6] . HRV, computed as the variability of the inter-beat interval (RR) series, measures the autonomic dysfunction caused by COPD [8], [9]. HRV has traditionally been quantified using time- and/or frequency-domain features computed from the RR time series [10]. Additionally, the existence of myocardial infarction as a co-morbidity with COPD makes it an important modality for measurement [3]. HRV measurements have traditionally been conducted on short term 5-minute RR time series for various applications [10]. However, shorter duration segments have been used in other wearable applications [11].

However, these studies have been conducted in controlled laboratory environments [6] using instantaneous RR information available from the unprocessed physiological signals [6], [7]. Further, the experiments relied on presence of a physician when collected in hospital settings [7]. Such data collection methods often try to remove effects of activity such as movement artefacts, light and temperature, the circadian cycle, and other confounding variables [12], as well as experimental noise during data collection. Hence, when used "in-the-wild" these methods perform very poorly, if at all [13]. Additionally, instantaneous HR might not be available for wearable devices for long-term monitoring. Here, HRV calculated from low-resolution HR data provided by commercially available devices is used and may not be comparable to raw signals, thus further decreasing performance [14].

In this work, we propose the use of smartwatch data collected in-the-wild for prediction of COPD severity levels and exacerbation. We show that a combination of activity and HRV features give the best performance by using the first few days of each subject data for training. To our knowledge, this is the first long-term study conducted in an in-the-wild setting showing the potential for remotely predicting COPD symptoms/severity using commercially available smartwatches.

II. Materials and Methods

A total of 35 (18 females, mean age: 69.9 $\pm$ 9.05) participants with COPD were recruited for a three month,

TABLE I

LIST OF SYMPTOMS IN THE DAILY QUESTIONNAIRE

Major	worsening breathlessness, change in sputum colour, increased sputum amount
Minor	a cold (runny or blocked nose), increased wheeze or chest tightness, sore throat, worsening cough

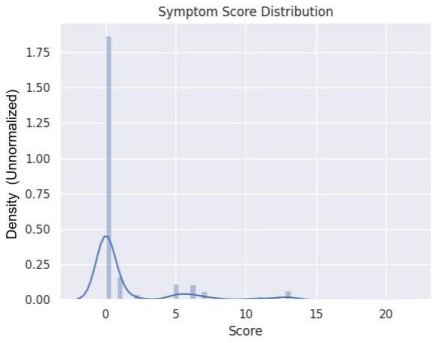

Fig. 1. Distribution of symptom scores across all participants

ethics review board (ERB) approved study. The participants were given a smartwatch and a smartphone with the data collection application pre-installed. Participants first gave informed consent regarding the type of data being collected and were instructed about the in-built privacy features. The participants were then asked to wear the watch during their daily activities and charge the watch at night. However, eight participants dropped out early on due to either personal reasons or difficultly in using the data collection application. The dropped out set still consented for the use of their data.

The participants filled out a daily questionnaire regarding their health using a smartphone application. The symptom questionnaire, scoring system, and exacerbation definition are adapted from [15]. They were asked to select symptoms that were "worse than usual" from a list of given symptoms (Table I). Each symptom is weighted as 1 (minor) or 5 (major), and the sum of weights is the daily total *symptoms score*; Fig. 1 gives the distribution of the symptoms scores. Additionally, an *exacerbation* is defined as two consecutive days with symptoms score of 6 or above – indicating two consecutive days with at least one major and one minor symptom.

A. Wearable Sensors

Each participant was given an Android Wear Smartwatch (Samsung Galaxy Watch, 42mm or 46mm) and a Samsung Note 9 smartphone. The smartwatch records audio data (from microphone), activity (from gyroscope and accelerometer), and HR (from photoplethysmography, PPG, at a sample rate of 100 Hz) information. To preserve the battery life of the device, we use a 20% duty cycle: 2 minutes continuous collection followed by

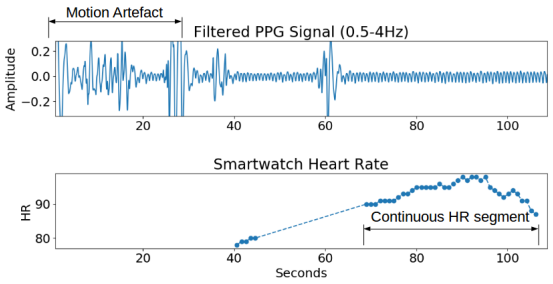

Fig. 2. Representative PPG and HR provided by the smartwatch

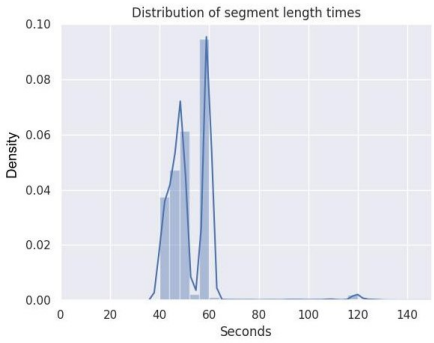

Fig. 3. Distribution of lengths of usable HR segments

8 minutes of inactivity. During the collection periods, the smartwatch provides HR values every second. The device also contains an internal quality metric for PPG signals and when segments are deemed "too noisy," the HR data is not recorded. Activity information is provided by the watch as total steps and the duration for the activity along with the hour in which the activity was recorded. For this paper, we only focus HR and activity information provided by the smartwatch.

B. Pre-processing

PPG data collected in-the-wild can have added noise due to several sources, including motion artifacts, tightness of watch, and environmental light [16]. Figure 2 shows an example PPG signal (bandpass filtered: 0.4-5Hz) with clean PPG and motion artifact regions captured by the smartwatch, along with the heart rate provided by the watch's internal PPG detection algorithm. As a result, the HR information provided by the smartwatch may have missing values, which can greatly impact the calculation of the HRV metric [17]. In order to ensure the quality of the HRV metric calculated, segments with more than 10 continuous seconds of missing HR information are discarded as unusable.

Fig. 3 shows the distribution of lengths of usable segments. Most of the usable segments are between 40-60 seconds in length and only a few segments being a full 2 minutes of PPG recording. Short-term HRV analysis has been generally conducted with 5 minute

TABLE II
BENCHMARK HRV FEATURES EXTRACTED

Type	Features
Time	meanRR, standard deviation (sdRR), coefficient of variation, rmsdd, pNN50, pNN20
Frequency	High frequency power (HF), normalized HF, Low frequency power (LF), normalized LF, very low frequency power, total power, LF/HF

windows, as recommended by [10]. However, smaller segments referred to as ultra-short term HRV have been tested in various domains [11]. We only considered HR segments with lengths $> 45s$ for HRV calculation in order to preserve the low frequency (0.04-0.15 Hz) component of the HRV signal [11]. Before HRV feature calculation, the HR segments were converted to RR values and a range-based filter was applied to remove any outliers. The range-based filter removes physiologically impossible values detected by the smartwatch algorithm by removing values below 350ms and above 1200ms.

C. Feature Extraction

1) HRV Features: We extract standard time- and frequency-domain HRV metrics; a complete list can be found in Table II. A majority of these features have been shown in the literature to correlate with COPD severity levels [6], [7]. Complete details about these measures can be found in [10].

HRV features were first extracted for each segment. Next, features were aggregated over a window of the last 12 or 24 hours before the survey was answered, to capture the long duration variability of HRV. The aggregation was done using the following statistical functionals: mean, standard deviation, min, max, skewness, and kurtosis. This resulted in a total of 78 (13×6) HRV features.

2) Activity Features: Physical activity is among the strongest predictor of mortality in COPD patients [18] and increased activity is known to improve COPD symptoms [19]. We used activity data provided by the smartwatch to derive several activity features to measure this change. These features were calculated based on the activity levels in the 12- or 24-hour window before survey was answered. The features correspond to the total steps taken, the total activity time, number of active hours, and number of inactive hours, all computed over the analyzed window. A total of 82 features (78 aggregated HRV and 4 activity features) per 12h or 24h window are available for prediction.

D. Prediction Pipelines

1) Symptom Score Prediction: Here, we made use of a past-future model for evaluation. The models were trained on the first X days (X = 7 to 14) of data for the all subjects as the training set. The next 7 days of data (from day X+1 to day X+7) from each subject was used as the validation set for parameter tuning. The remaining days were used for the test set. Subjects were only included in the analysis if they had at least four weeks of data. As a result, we evaluated 8 and 12 subjects for the 12- and 24-hour feature aggregation settings, respectively.

Daily COPD symptoms scores were binarized using a threshold of 3 into low and high severity levels and a binary classification was performed on the labels. Hence, a score with all four of the minor symptoms or at least one major symptom are categorized as high severity while anything less than that are considered low severity. The models have a number of hyper-parameters that need to be tuned. Optimal hyper-parameters were searched on the validation dataset and used to make a final predictions on the test set. We report these final predictions as our result. Parameters searched include:

1) number of first X days (X = 7 to 14) to use for each subject;
2) HRV features alone or combined HRV-activity;
3) feature set aggregation window: 12 and 24 hours;
4) classifier: SVM (with RBF kernal, class weight: balanced), logistic regression, and a random forest (with 20 trees). Default values were used for all other parameters of the classifiers;
5) sampling, where two oversampling methods were tested, including the standard oversampling strategy and synthetic minority oversampling technique (SMOTE). SMOTE is a data augmentation method that creates new samples for the minority class drawing new samples from the linear interpolation between neighboring samples [20].

Training traditional machine learning classifiers with a large number of features may lead to overfitting and many features may exhibit high correlations amongst them, thus not contributing to the prediction task. As such, recursive feature elimination was performed with a step size of 1 using the extra trees classifier for feature ranking. Here, the top 10 features were selected before prediction for each hyper-parameter setting. Due the dataset imbalance ($\%$ imbalance: 0.19), balanced accuracy (BACC), F1-score (F1) and Matthews correlation coefficient (MCC) are used as figures-of-merit. An MCC value of 0 represents random prediction while 1 represents perfect prediction and has been shown to be a robust metric for imbalanced data [21]. Additionally, the receiver operating characteristic (ROC) curve along with the AUC value are also shown for the optimal setting. The performance was also compared to a baseline classifier that randomly chooses based on the class distribution in the training test (called a random choice classifier). The hyper-parameters were searched on the validation

TABLE III
TOTAL DATA SUMMARY FOR 12- AND 24-HOUR AGGREGATION SETTINGS

Hours	Subjects	Samples	%Severe	%Exacerbation
12	8	862	24%	15%
24	12	1393	22%	14%

dataset and optimal values chosen based on maximizing the MCC metric.

2) Exacerbation Prediction: For prediction of exacerbation, an anomaly detection approach was used with the exacerbation events treated as anomalies due to the small number of such events in the dataset. An isolation forest [22] with default parameters was used as the classifier to train the normal samples of the training data. For isolation forest, the measure of normality of an observation given a tree is the depth of the leaf containing this observation, which is equivalent to the number of splittings required to isolate this point in the feature space. An anomalous point lies closer to the root of the tree resulting in smaller depth of the leaf. Hence, the smaller the measure of normality the more anomalous a given point is. Due to the small number of exacerbation samples in both the training and validation sets, no hyper-parameter search was performed in this case. The optimal hyper-parameters selected for severity prediction were used for prediction of exacerbation. The same performance metrics used for severity prediction have been used for this along with performance comparison to a random choice classifier. The pipelines were implemented using the scikit-learn [23] and imbalanced-learn [20] libraries. Table III shows a summary of the data and percentage of positive labels for severity and exacerbation prediction.

III. EXPERIMENTAL RESULTS AND DISCUSSION

A. Severity prediction

Table IV shows the results for the proposed severity predictor along with a random choice classifier. Figure 4, in turn, shows the corresponding ROC curve. We observe that the performance is better than chance level for all metrics. The 24-hour aggregation models, combination of both HRV and activity features, SMOTE oversampling strategy, random forest classifier, and using X=9 initial days of data were the optimal hyperparameters found. Overall, we see improvements of 0.15 in MCC, 0.09 in BACC and 0.08 in F1 over the random choice baseline.

An analysis of the top-10 features selected showed that two functionals of meanRR appear in the top feature set. A decreased meanRR is an indicator of sympathetic dominance caused by COPD [24]. Also

TABLE IV
PERFORMANCE ON SEVERITY PREDICTION

	BACC	F1	MCC
Proposed	0.61	0.41	0.19
Random Choice	0.52	0.33	0.04

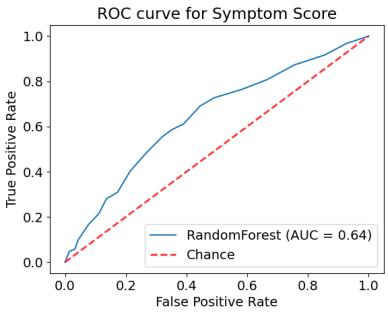

Fig. 4. ROC curve for severity prediction

sdRR and coefficient of variation appear as top features. They have been used as an indicator of improvements in COPD condition with physical activity [19]. A decreased sdRR may also suggest a decrease in HRV indicating autonomic dysfunction [9]. Both LF and HF power are in the top feature set and are also indicative of autonomic dysfunction [9] in COPD. Number of active hours appears as a top feature showing the importance of physical activity on COPD levels [18], [19]. Overall, 9 of the top 10 features are HRV related showing its overall importance in detection of COPD severity levels.

B. Exacerbation prediction

Table V shows the results for the proposed exacerbation prediction along with the random choice classifier performance. Figure 5, in turn, shows the corresponding ROC curve using the anomaly scores output from the isolation forest. Again, we observe that performance is better than chance for all metrics. Overall, we see improvements of 0.12 in MCC, 0.08 in BACC and 0.04 in F1 over the random classifier.

One limitation of the proposed method is that it assumes availability of initial calibration data (X) for each subject in order to train the models. A solution to this would be a model trained with data from other subjects only leading to a leave-one-subject-out (LOSO) evaluation strategy. However, the small amount

TABLE V
PERFORMANCE ON EXACERBATION PREDICTION

	BACC	F1	MCC
Severity	0.57	0.27	0.12
Random Choice	0.49	0.23	0.0

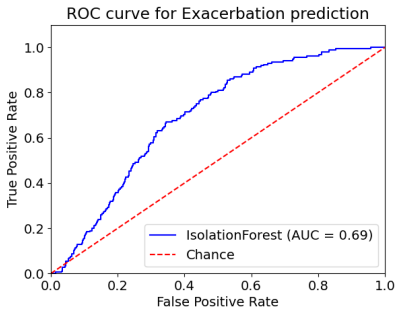

Fig. 5. ROC curve for exacerbation prediction

of subjects limits the generalizability and reliability of a LOSO approach. Moreover, due to the small number of exacerbation events, no hyperparameter tuning was performed for exacerbation prediction. While the results obtained using the hyperparameters from severity prediction outperform random choice classifier, a larger database can allow for further improvements via hyperparameter tuning.

IV. Conclusion

In this paper, we evaluate COPD severity and exacerbation using HR and activity data collected remotely from subjects using commercially available smartwatches during their daily life. We show that future trajectory can be predicted with the use of only a few days of subject data for calibration of models. Symptoms can also be predicted with accuracies above chance with a combination of HRV and activity features. Overall, the results suggest that COPD patient monitoring may be performed remotely, thus reducing the burden on an already overwhelmed healthcare system.

References

[1] P. Burney, J. Patel, R. Newson, C. Minelli, and M. Naghavi, "Global and regional trends in COPD mortality: 1990–2010," *European Respiratory J.*, vol. 45, no. 5, pp. 1239–1247, 2015.

[2] H. Svedsater, J. Roberts, C. Patel, J. Macey, E. Hilton, and L. Bradshaw, "Life impact and treatment preferences of individuals with asthma and chronic obstructive pulmonary disease: results from qualitative interviews and focus groups," *Advances in therapy*, vol. 34, no. 6, pp. 1466–1481, 2017.

[3] I. Tomasic, N. Tomasic, R. Trobec, M. Krpan, and T. Kelava, "Continuous remote monitoring of copd patients—justification and explanation of the requirements and a survey of the available technologies," *Medical & biological engineering & computing*, vol. 56, no. 4, pp. 547–569, 2018.

[4] N. Liu, R. Huang, T. Baldacchino, A. Sud, K. Sud, M. Khadra, and J. Kim, "Telehealth for noncritical patients with chronic diseases during the covid-19 pandemic," *Journal of Medical Internet Research*, vol. 22, no. 8, p. e19493, 2020.

[5] J. Peake, G. Kerr, and J. Sullivan, "A critical review of consumer wearables, mobile applications, and equipment for providing biofeedback, monitoring stress, and sleep in physically active populations," *Frontiers in physiology*, vol. 9, p. 743, 2018.

[6] M. Rahman, E. Nemati, M. Rahman, V. Nathan, K. Vatanparvar, and J. Kuang, "Automated assessment of pulmonary patients using heart rate variability from everyday wearables," *Smart Health*, vol. 15, p. 100081, 2020.

[7] C. Bellos, A. Papadopoulos, R. Rosso, and D. Fotiadis, "Identification of copd patients' health status using an intelligent system in the chronious wearable platform," *IEEE journal of biomedical and health informatics*, vol. 18, no. 3, pp. 731–738, 2013.

[8] A. Bernardo, L. Vanderlei, and D. Garner, "HRV analysis: A clinical and diagnostic tool in chronic obstructive pulmonary disease," *International scholarly research notices*, 2014.

[9] A. Gestel and J. Steier, "Autonomic dysfunction in patients with chronic obstructive pulmonary disease (copd)," *Journal of thoracic disease*, vol. 2, no. 4, p. 215, 2010.

[10] A. Camm *et al.*, "Heart rate variability: standards of measurement, physiological interpretation and clinical use. task force of the european society of cardiology and the north american society of pacing and electrophysiology," *Circulation*, vol. 93, no. 5, pp. 1043–1065, 1996.

[11] A. Tiwari, R. Cassani, J.-F. Gagnon, D. Lafond, S. Tremblay, and T. Falk, "Prediction of stress and mental workload during police academy training using ultra-short-term heart rate variability and breathing analysis," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 2020, pp. 4530–4533.

[12] I. Ray, D. Liaqat, M. Gabel, and E. de Lara, "Skin tone, confidence, and data quality of heart rate sensing in wearos smartwatches," in *2021 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, 2021.

[13] D. Liaqat, R. Wu, S. Liaqat, E. de Lara, A. Gershon, and F. Rudzicz, "The ground truth trade-off in wearable sensing studies," *arXiv preprint arXiv:2001.09738*, 2020.

[14] A. Tiwari, R. Cassani, S. Narayanan, and T. Falk, "A comparative study of stress and anxiety estimation in ecological settings using a smart-shirt and a smart-bracelet," in *International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2019, pp. 2213–2216.

[15] S. Aaron, G. Donaldson, G. Whitmore, J. Hurst, T. Ramsay, and J. Wedzicha, "Time course and pattern of copd exacerbation onset," *Thorax*, vol. 67, no. 3, pp. 238–243, 2012.

[16] D. Castaneda, A. Esparza, M. Ghamari, C. Soltanpur, and H. Nazeran, "A review on wearable PPG sensors and their potential future applications in health care," *Intl. J. Biosensors & Bioelectronics*, vol. 4, no. 4, p. 195, 2018.

[17] M. Peltola, "Role of editing of rr intervals in the analysis of heart rate variability," *Frontiers in physiology*, vol. 3, p. 148, 2012.

[18] B. Waschki, A. Kirsten, O. Holz, K.-C. Müller, T. Meyer, H. Watz, and H. Magnussen, "Physical activity is the strongest predictor of all-cause mortality in patients with copd: a prospective cohort study," *Chest*, vol. 140, no. 2, pp. 331–342, 2011.

[19] C. Camillo, V. Laburu, N. Gonçalves *et al.*, "Improvement of heart rate variability after exercise training and its predictors in COPD," *Resp. medicine*, vol. 105, no. 7, pp. 1054–1062, 2011.

[20] G. Lemaître, F. Nogueira, and C. Aridas, "Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning," *The Journal of Machine Learning Research*, vol. 18, no. 1, pp. 559–563, 2017.

[21] S. Boughorbel, F. Jarray, and M. El-Anbari, "Optimal classifier for imbalanced data using matthews correlation coefficient metric," *PloS one*, vol. 12, no. 6, p. e0177678, 2017.

[22] F. Liu, K. Ting, and Z.-H. Zhou, "Isolation forest," in *Intl. Conf. on Data Mining*. IEEE, 2008, pp. 413–422.

[23] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg *et al.*, "Scikit-learn: Machine learning in python," *Journal machine learning research*, vol. 12, no. Oct, pp. 2825–2830, 2011.

[24] S. K. Chhabra, M. Gupta, S. Ramaswamy, D. J. Dash, V. Bansal, and K. Deepak, "Cardiac sympathetic dominance and systemic inflammation in copd," *COPD: Journal of Chronic Obstructive Pulmonary Disease*, vol. 12, no. 5, pp. 552–559, 2015.