Classification of Exacerbation Episodes in Chronic Obstructive Pulmonary Disease Patients

A. Dias\textsuperscript{1,2}; L. Gorzelniak\textsuperscript{2,3}; K. Schultz\textsuperscript{4}; M. Wittmann\textsuperscript{5}; J. Rudnik\textsuperscript{4}; R. Jörres\textsuperscript{5}; A. Horsch\textsuperscript{2,6}

\textsuperscript{1}Computer Science Department, University of Tromsø, Tromsø, Norway; \textsuperscript{2}Institute for Medical Statistics and Epidemiology (IMSE), Technische Universität München, Munich, Germany; \textsuperscript{3}Institute for Epidemiology, Helmholtz-Zentrum München, German Research Center for Environmental Health, Munich, Germany; \textsuperscript{4}Clinic Bad Reichenhall, Center for Rehabilitation, Pneumology and Orthopedics, Bad Reichenhall, Germany; \textsuperscript{5}Institute and Outpatient Clinic for Occupational, Social and Environmental Medicine, LMU, Munich, Germany; \textsuperscript{6}Department of Computer Science and Department of Clinical Medicine, University of Tromsø, Tromsø, Norway

Keywords
COPD, exacerbation, accelerometer, classification

Summary

Background: Chronic obstructive pulmonary disease (COPD) is a progressive disease affecting the airways, which constitutes a major cause of chronic morbidity and a significant economic and social burden throughout the world. Despite the fact that in COPD patients exacerbations are common acute events causing significant and often fatal worsening of symptoms, an accurate prognostication continues to be difficult.

Objectives: To build computational models capable of distinguishing between normal life days from exacerbation days in COPD patients, based on physical activity measured by accelerometers.

Methods: We recruited 58 patients suffering from COPD and measured their physical activity with accelerometers for 30 days. We were able to analyse data for 52 patients (369 patient days), and extracted three distinct sets of features from the data, one set of basic features such as average, one set based on the frequency domain and the last exploring the cross-information among sensors pairs. These were used by three machine-learning techniques (logarithmic regression, neural networks, support vector machines) to distinguish days with exacerbation events from normal days.

Results: The support vector machine classifier achieved an AUC of 90% ± 7, when supplied with a set of features resulting from sequential feature selection method. Neural networks achieved an AUC of 83% ± 16 and the logarithmic regression an AUC of 67% ± 15.

Conclusions: None of the individual feature sets provided robust for reasonable classification of PA recording days. Our results indicate that this approach has the potential to extract useful information for, but are not robust enough for medical application of the system.

1. Introduction

Chronic obstructive pulmonary disease (COPD) is a progressive disease affecting the airways. It constitutes a major cause of chronic morbidity and mortality and constitutes a considerable economic and social burden throughout the world [1]. It is the fifth leading cause of death worldwide, with forecasts indicating an increase of prevalence and mortality in the coming decades [2].

Despite exacerbations of COPD, which are defined as acute events with a significant worsening of lung function and symptoms, being both common and often fatal, it is difficult to accurately forecast the need of hospitalization for patients with an exacerbation. Patients with COPD experience an exacerbation about twice a year [3].

The mechanisms of COPD exacerbation are complex, with respiratory viruses and bacteria playing a major role in the cause of these events [4]. Patients who rapidly receive treatment after onset of symptoms have better outcomes than patients who wait several days to seek treatment.

Physical activity (PA) is an important measure in COPD, and PA decrease is currently hypothesized to be an early indicator of disease progression [5]. PA has been associated with lung function decline [6]. Accelerometers are an important tool for objective PA measurement in COPD [7].

Several studies [8–11] have investigated the forecasting of exacerbation episodes in patients with COPD but with only partial success. Marin et al. and Ong et al. pro-
posed to use the BODE index, a composite marker of COPD disease, to predict the probability of exacerbation within a year period. Miniati et al. also aimed at over the year prediction but using the C-reactive protein as source of information.

Proposing an innovative approach Jensen et al. [12] used data reflecting physiological parameters, recorded in a telecare project, and attempted to predict COPD exacerbations. They obtained encouraging results with an area under the curve (AUC) of 73%.

Other authors have explored the predictive capacity of PA in COPD, but aiming at the mortality rate [13] or the overall health status of the patients [14]. Garcia-Rio et al. found that the time until first admission to hospital due to exacerbation was linked to lower physical activity levels, which hints at a relation between PA and exacerbations.

Researchers have explored the associations of data from accelerometers to other medical conditions or outcomes, such as the risk of fall in elderly people and gait parameters. Marschollek et al. [15] proposed an accelerometer-based system to assess the risk of fall in people above 65 years old, an age group with higher risk and with serious consequences of falls. They achieved fair to good classification performance. Their work provides some of the inspiration for our research. Gietzelt et al. [16] have also used accelerometers as a tool for measurement of gait parameters in people with dementia, achieving good accuracy in the most relevant parameters.

To our knowledge there are no studies, yet, that explore the use of continuous PA measurements to predict exacerbations, regardless of the timeframe. This work explores the ability to distinguish the exacerbation episodes from normal days by using the rolling PA measurements that precede the cases in time. This is a first insight to the possibility of forecasting COPD exacerbations by a simple method. If this proves to be feasible, industry may in the future be able to develop a wearable device with an accelerometer and on-board data processing that can alert the COPD patient some time before the onset of an exacerbation. It is even conceivable that patients may be alerted about an impending exacerbation episode, i.e., before symptoms are developed fully.

2. Material and Methods

2.1 Study Subjects and Protocol

We recruited 58 patients with COPD (35 males, 23 females) undergoing LTOT who had been admitted for a 3-week in-patient Pulmonary Rehabilitation (PR) program, in the scope of another study [17], from August 2009 to March 2010. The diagnosed COPD in all cases was related to smoking at stage IV of the GOLD classification. We processed the data for 52 patients, after we excluded the data for 6 of them (5 males, 1 female), either because they didn’t actually wear the sensors (non-compliance, 2) or because of technical sensor failures (4). We obtained an informed consent for all participants; the study protocol was approved by the local ethics committee and conducted according to the Declaration of Helsinki.

We asked each participant to wear three accelerometers for at least 10 days, placed at hip, ankle and wrist since getting up until going to bed. During this period patients were undergoing the PR program, which consists of daily exercise training and physiotherapy as well of breathing exercises.

2.2 Accelerometers

We assessed the PA of wrist and ankle using the uniaxial piezoelectric GT1M accelerometer (Actigraph LLC, Pensacola FL, firmware version 4.2.0), attached by means of an elastic belt and labeled in order to avoid misplacement. The GT1M is a small, lightweight system that measures acceleration in the vertical plane and expresses PA as “activity counts”, a proprietary quantification of the detected accelerations, in time intervals of 1 minute [18]. The Actigraph is a validated device and the most widely used device in research [19].

Assessment of the hip movements was performed by the triaxial RT3 accelerometer (Stayhealthy, Monrovia, CA, firmware version 0.6), worn in a holster at the non-dominant side of the waist. The RT3 records activity of the three orthogonal directions as vector magnitude units (VMU) in time intervals of 1 minute [20]. A previous version of this sensor has been used successfully in patients with COPD [21]. We synchronized the clock on each sensor with the clock of the same computer, which ensured synchronicity of the three sensors.

All subjects were carefully instructed about how to position the devices, for half an hour (Figure 1), and they received a manual with clear instructions and illustrative diagrams.

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Baseline characteristics of the patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>Characteristics</td>
<td>All patients (N = 52)</td>
</tr>
<tr>
<td>Age, years</td>
<td>62 ± 10</td>
</tr>
<tr>
<td>Gender (F/M)</td>
<td>22/30</td>
</tr>
<tr>
<td>BMI (kg/m2)</td>
<td>26 ± 7</td>
</tr>
<tr>
<td>FEV1 (m)</td>
<td>265 ± 94 (N = 37)</td>
</tr>
<tr>
<td>Dyspnea Scale MMRC</td>
<td>3 ± 1 (N = 40)</td>
</tr>
<tr>
<td>Borg CR 10</td>
<td>5 ± 2 (N = 36)</td>
</tr>
<tr>
<td>FEV1 %VC</td>
<td>45 ± 13</td>
</tr>
</tbody>
</table>

Data expressed as mean ± SD. BMI, Body Mass Index; PR, pulmonary rehabilitation; MMRC, modified Medical Research Council (MMRC) dyspnea scale; 6MWD, 6-Minute Walk Distance; Borg CR 10, patients’ self-perceived degree of breathlessness at the end of the 6MWT using the modified Borg CR 10 scale; FEV1, forced expired volume in 1 s; VC, vital capacity. Not all patients conducted 6MWD, MMRC or Borg CR, for medical reasons.

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Characteristics and time frame for COPD exacerbation in the six patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)/ Gender</td>
<td>Day of onset</td>
</tr>
<tr>
<td>76/m</td>
<td>2</td>
</tr>
<tr>
<td>72/m</td>
<td>12</td>
</tr>
<tr>
<td>64/m</td>
<td>3</td>
</tr>
<tr>
<td>50/f</td>
<td>2</td>
</tr>
<tr>
<td>56/m</td>
<td>5</td>
</tr>
<tr>
<td>58/f</td>
<td>3</td>
</tr>
</tbody>
</table>

Day of onset is expressed as days since start of accelerometer monitoring. Duration indicates the number of days in the exacerbation episode for which we had accelerometer data.
2.3 Exacerbation Episodes

The medical staff of the clinic collected the date of onset and recovery of exacerbation episodes from the patients’ medical record. An exacerbation was defined if the patient was treated with oral corticosteroids in combination with or without antibiotics. For all patients we only have recorded onset of episodes during the accelerometer measurement period. All patients recovered only after we stopped the assessment; thus, no data on the recovery phase of the exacerbation was available for analysis. We recorded six exacerbation episodes in six different patients. The onset of the exacerbation ranged from the third day to the twelfth day after start of physical activity recording.

2.4 Feature Extraction

In an initial approach to predict exacerbation episodes we focused on a simpler task of classifying exacerbation days from control days. For each day of the study, \( n \), we extracted features from the data set of day \( n - w, \ldots, n - 1 \) and \( n \). We did not consider the day of diagnosis for feature extraction, as the medical status was unclearly defined as normal or exacerbation day. With \( w \) being a feature windowing indicating how many days before the \( n \) day we considered the data for feature extraction. We considered those data from the day before diagnosis of the exacerbation can be already influenced by the onset and, so cannot be safely considered a non-exacerbation. We built 3 sets of features from the accelerometer data for classification. The first set (Set 1) quantifies the characteristics of the PA data from each individual sensor separately. These are standard features used in machine learning, independent of the application domain, extracting basic information from the time series. The second set (Set 2) included frequency and time-scale features from each sensor. These are features commonly used in signal processing problems, and have been previously used in the classification of accelerometer data [22], namely in the scope of activity detection through accelerometers. They are based on the frequency domain analysis of the data, an important tool for signal processing. In the last set (Set 3) we explored the cross-information of data collected at different body parts, as the placement of the sensors influences the type of information recorded. For this we implemented features based on the cross information among sensor pairs, and thus body parts. We also explored a set of features composed of all the features together for which we selected the most important using a serial feature selection method (Set 4).

Set 1 was composed of the features (one value for each one of the three sensors):

- Mean (\( \mu \)), standard deviation (SD), skewness (SK), kurtosis (KT), maximum (MAX), minimum (MIN) and linear correlation coefficient (LC): we used the common implementations of these functions over the data of the corresponding days.
- Mean crossing rate (MC) – Total number of times the signal crosses the mean value of the whole period. Provides an estimation of the movement pattern, specifically for signals that are subject to noise.
- Autocorrelation over 24 hours (AC24) – Indicates the similarity of the data over a period of 24 hours to the next period. It is an estimation of the changes of activity from day to day.
- 24 hours autocorrelation after 5 minutes moving average (AC245) – Similar feature as the previous one but after the original data has been smoothed with a moving average filter with 5 minutes window.

Set 2 was composed by the features in the frequency and time-scale domain; all features were extracted after applying a fast Fourier transformation (each sensor):

- Mean frequency (MF) – Calculates the average frequency of the movement in the period.
- Dominant frequency (DF) – Calculates the most common frequency in the movements. It indicates the nature of the movements in terms of frequency.
- Energy (E) – Calculates the energy, normalized over each sub-band, of the spectrum of the data.
- Linear correlation coefficient (LCF) – The frequency domain linear correlation coefficients reveal the presence of a strong correlation (between 0.6 and 0.99) for similar movements.

Set 3:

- Cross-correlation of the data from the sensor pairs hip-leg (HLC), hip-arm (HAC) and leg-arm (LAC) – This feature indicates the extent of the offset, if any, between the two data series from different sensors.
- Correlation coefficient of hip-leg (HLCC), hip-arm (HACC), leg-arm (LACC) – It is an estimation of the similarity between the data from two different sensors, indicating the coordination of movements between different body parts.
For all the above features we calculated them twice with time window, \( w \), of two and three days before the feature day. We explored the best time window by running tests with window values ranging from one to five days (half the recorded period). Given the time frames of the exacerbation episodes and the amount of data recorded before the onset event, we found that smaller windows would not carry significant information and larger values would reduce the number of possible features too much. We have, for instance the feature \( \text{MAX}_2 \) and \( \text{MAX}_3 \), respectively the maximum value calculated over the data series of two days and three days the feature day.

In addition to the three sets of features we further merged all features into a large feature set and selected the most relevant ones by applying the sequential feature selection algorithm. Of the original 90 features, the algorithm selected 17 and we ran the classification methods based on them. The selected features were: \( \text{SD}_2 \), \( \text{SK}_2 \), \( \text{LC}_2 \), \( \text{MC}_3 \), \( \text{AC24}_2 \), \( \text{AC24}_3 \), \( \text{MF}_2 \), \( \text{MF}_3 \), \( \text{E}_2 \), \( \text{E}_3 \), \( \text{LCF}_2 \), \( \text{LCF}_3 \), \( \text{HLC}_3 \), \( \text{HACC}_2 \), \( \text{HACC}_3 \), \( \text{HLCC}_2 \), \( \text{HLCC}_3 \). The numbers indicate the time window \( w \). We did not reduce further the number of features, which could compromise the overall performance.

### 2.5 Classification Algorithms

After extracting all the features we implemented three different classification methods to apply to the datasets: 1) logarithmic regression (LOG), 2) support vector machines (SVM), and 3) feed-forward neural network (NN) with 50 neurons in the hidden layer. The configuration for the neural network was established after running a test with 5, 10, 20, 50 and 100 neurons and finding the most robust one. We used a 10-fold cross validation method for assessing the robustness of the algorithms, given the low number of test and control cases in the dataset. For each algorithm and feature set the ROC curve is reported as an average of the ten runs. Both the feature extraction and classification were implemented in Matlab version 7.0.4 (R14SP2) running on GNU Linux (Ubuntu 10.4 and 12.04).

### 3. Results

#### 3.1 Set 1

Figure 2 displays the ROC for the three classification algorithms when using set 1 of features. The support vector machine has the best performance with an AUC of 74% (Table 1). The logarithmic regression algorithm has a very poor performance, with negative classification power.

![Figure 2](image1.png)

**Figure 2**

ROC curve for feature set 1

![Figure 3](image2.png)

**Figure 3**

ROC for the three classification algorithms over feature set 2
Overall set 1 did not provide useful classification information.

### 3.2 Set 2

Figure 3 shows the ROC of the set 2 of features, which shows an improvement of the classification power of the features and methods over set 1. The SVM achieved the best results as well, with NN unable to improve the AUC with the new feature set.

### 3.3 Set 3

Figure 4 shows the performance of the three algorithms for the set 3 of features. The SVM classifier was the best classifier for this feature set. The logarithmic regression and NN achieved comparable AUC.

### 3.4 All Features

Figure 5 displays the ROC for the features after applying the sequential feature selection method to a merge of all previous feature sets, which resulted in an increase in the classification power of the three methods. Over these features SVM was the best method, achieving an AUC of 90%. NN in particular showed a further increase of the AUC for this set, over the previous partial feature set.

### 3.5 Comparison

Table 3 shows the AUC for the four feature sets and the three classifiers. Overall SVM achieved the highest AUC while logarithmic regression showed the lowest performance. Overall the results indicate a fairly low classification capacity of the different algorithms. With the best classifier SVM, using all the features, at a specificity level of 85% we can achieve 100% sensitivity.

### 4. Discussion

This study is, to our knowledge, the first attempt to extract a set of features from accelerometer data to classify exacerbation episodes in COPD. Using the best feature set we achieved in classifying exacerbation episodes an AUC of 67% with logarithmic regression, 83% with NN and 90% with SVM.

We carried the study in an in-clinic setting, with patients under daily medical supervision. Thus the generalization of the results to a daily living setting is limited.

---

**Table 3**

<table>
<thead>
<tr>
<th>Features</th>
<th>Algorithm (AUC and standard deviation)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Logarithmic regression</td>
</tr>
<tr>
<td>Set 1</td>
<td>45.0% (13.6%)</td>
</tr>
<tr>
<td>Set 2</td>
<td>65.7% (14.0%)</td>
</tr>
<tr>
<td>Set 3</td>
<td>58.9 (12.1)</td>
</tr>
<tr>
<td>All</td>
<td>66.5 (15.3)</td>
</tr>
</tbody>
</table>
Overall the classifiers achieved poor performance, preventing the application of the method to a medical setup, even if the alternative is having no classification method at all. Even if the SVM had a 90% AUC for the last set, the other classifiers never exceeded 82%.

Overall SVM showed to be the most robust classifier for this task. As for features, none of the basic feature sets, with distinct types of features, proved robust enough for this problem. Only the combination of all the extracted features, after a feature selection process, provided enough information for reasonable classification power.

The findings indicate that sensors can provide relevant information for the task of assessment of the patient state; however, the extraction process was not optimized and produced results that are not amenable for clinical applications.

To further improve this approach and potentially provide results able to assist the patients on managing the disease at home we need to identify better features to extract relevant information from the accelerometer data and conduct studies that record a significant number of exacerbations in a daily living setting. The quality of the results may also improve by use of intra-individual models, where each algorithm is trained only with data of one patient. But for this approach to be feasible long recording periods will be needed, comprising several weeks or months of data and comprising several exacerbation events.

In our study we used a very simple approach for feature extraction. We did not experiment with existing algorithms to identify activities from the accelerometer data and conduct studies that record a significant number of exacerbations in a daily living setting. The quality of the results may also improve by use of intra-individual models, where each algorithm is trained only with data of one patient. But for this approach to be feasible long recording periods will be needed, comprising several weeks or months of data and comprising several exacerbation events.

In our study we used a very simple approach for feature extraction. We did not experiment with existing algorithms to identify activities from the accelerometer data and conduct studies that record a significant number of exacerbations in a daily living setting. The quality of the results may also improve by use of intra-individual models, where each algorithm is trained only with data of one patient. But for this approach to be feasible long recording periods will be needed, comprising several weeks or months of data and comprising several exacerbation events.

In our study we used a very simple approach for feature extraction. We did not experiment with existing algorithms to identify activities from the accelerometer data and conduct studies that record a significant number of exacerbations in a daily living setting. The quality of the results may also improve by use of intra-individual models, where each algorithm is trained only with data of one patient. But for this approach to be feasible long recording periods will be needed, comprising several weeks or months of data and comprising several exacerbation events.

In our study we used a very simple approach for feature extraction. We did not experiment with existing algorithms to identify activities from the accelerometer data and conduct studies that record a significant number of exacerbations in a daily living setting. The quality of the results may also improve by use of intra-individual models, where each algorithm is trained only with data of one patient. But for this approach to be feasible long recording periods will be needed, comprising several weeks or months of data and comprising several exacerbation events.

In our study we used a very simple approach for feature extraction. We did not experiment with existing algorithms to identify activities from the accelerometer data and conduct studies that record a significant number of exacerbations in a daily living setting. The quality of the results may also improve by use of intra-individual models, where each algorithm is trained only with data of one patient. But for this approach to be feasible long recording periods will be needed, comprising several weeks or months of data and comprising several exacerbation events.

In our study we used a very simple approach for feature extraction. We did not experiment with existing algorithms to identify activities from the accelerometer data and conduct studies that record a significant number of exacerbations in a daily living setting. The quality of the results may also improve by use of intra-individual models, where each algorithm is trained only with data of one patient. But for this approach to be feasible long recording periods will be needed, comprising several weeks or months of data and comprising several exacerbation events.

In our study we used a very simple approach for feature extraction. We did not experiment with existing algorithms to identify activities from the accelerometer data and conduct studies that record a significant number of exacerbations in a daily living setting. The quality of the results may also improve by use of intra-individual models, where each algorithm is trained only with data of one patient. But for this approach to be feasible long recording periods will be needed, comprising several weeks or months of data and comprising several exacerbation events.

The main limitation of the study is the small data sample of exacerbation episodes we managed to acquire during the observation phase. Although we included a large number of patients in the study, we are dealing with unpredictable exacerbation occurrences making it hard to estimate and achieve a large number of episodes during a monitoring period. To achieve a large number of exacerbations we would need to ask the patients to wear the sensors for long periods (months) with all the usability and acceptance hindrances that would arise.

Acknowledgements
The authors thank the medical team in the rehabilitation clinic Bad Reichenhall, and the colleagues at Aalborg University Morten Jensen and Prof. Ole Hejlesen for discussion on the feature extraction and classification methods.

Funding
This research was funded/supported by the Graduate School of Information Science in Health (GSISH) and the Technische Universität München Graduate School. André Dias is supported by the Portuguese Foundation for Science and Technology (FCT), by scholarship SFRH/BD/39867/2007 and Research Council of Norway Grant No. 174934.

Contribution of the Authors
André Dias is the guarantor of the entire manuscript.
Lukas Gorzelniak contributed to the patient recruitment, data screening, analysis and the “Introduction” chapter.
Michael Wittmann, was involved in recruitment of patients, and a reviewer of the manuscript.
Konrad Schultz, was involved in recruitment of patients, a reviewer of the manuscript, provided references.
Rudolf Jörres discussed the study protocol, is the key reviewer of the entire manuscript, established the collaboration between both clinics and the researchers and gave feedback to all sections of this document.

Juliane Rudnik discussed the study protocol, is a key reviewer of the manuscript and provided the data regarding the exacerbation episodes.
Alexander Horsch discussed the study protocol, established the collaboration between both clinics and the researchers, is a key reviewer of the entire manuscript, and identified the study limitations.

List of Abbreviations
6MWD = 6-minute walk distance
BMI = Body Mass Index
FEV1 = forced expiratory volume in 1 s
LTOT = long-term oxygen therapy
MMRC = modified Medical Research Council dyspnea scale
PA = physical activity
PR = pulmonary rehabilitation
SD = standard deviation
VC = vital capacity
VMU = vector magnitude units

References