Early experiences in COPD exacerbation detection

Mario Merone*, Leonardo Onofri†, Paolo Soda‡, Claudio Pedone‡, Raffaele Antonelli Incalzi‡, Giulio Iannello*
*Computer Systems and Bioinformatics Laboratory, Università Campus Bio-Medico di Roma, Rome, Italy
†Dipartimento di Ingegneria Elettrica e delle Tecnologie dell’Informazione Universita’ degli Studi di Napoli Federico II, Napoli, Italy
‡Area di Geriatria, Università Campus Bio-Medico di Roma, Rome, Italy
{m.merone,l.onofri,p.soda,g.iannello}@unicampus.it

Abstract—Chronic obstructive pulmonary disease (COPD) is a slowly progressive disease characterized by airway obstruction. Patients suffering from COPD could report exacerbations, which are deteriorations in respiratory health that worsen the course of the disease. The prompt recognition of an exacerbation from daily variations and its early treatment reduces the healing time and the risk of hospitalization. Using a pulse oximeter connected to a mobile phone, we remotely collect the values of heart rate and of the oxygen saturation on a cohort of seven elderly patients affected by severe COPD. We analyze if a score given by the weighted composition of these signals permits to detect the worrisome events prefiguring an exacerbation onset. A cross validation-based evaluation allows us to assess how the results generalize to an independent data set. We found that the tested score does not provide satisfactory results in terms of sensitivity and specificity, suggesting that it is not able to disambiguate between exacerbation onset and other COPD related events.

I. INTRODUCTION

Chronic obstructive pulmonary disease (COPD) is a slowly progressive disease characterized by airway obstruction, producing a totally or partially reversible airflow limitation. It is caused by a chronic inflammation of both the airways and the lung parenchyma [1]. Recent research of the World Health Organization and of the Global initiative for chronic obstructive lung disease revealed that nowadays the COPD is the 4th cause of death in the United States coming next to heart disease, cancer, and cerebrovascular diseases. A similar statistic holds also for the entire world, where the COPD can be considered the 5th cause of death [2], thus emphasizing the relevance of this disease.

Patients suffering from COPD could report exacerbations, which are deteriorations in respiratory health that aggravate the course of the disease [3]. They are characterized by a worsening of symptoms and, caused also by respiratory viruses and bacteria, they should increase morbidity and mortality. Although patients with COPD experience day-to-day symptom variations, an exacerbation can be recognized by the appearance of a pattern of symptoms characterized by increased dyspnea, more productive cough than usual, which often becomes purulent sputum determining a deterioration in the quality of life and the need to change therapy [2]. Distinguishing the daily variations from exacerbations is challenging but important. Indeed, a prompt recognition of an exacerbation and its early treatment reduces the healing time and the risk of hospitalization, with a relevant impact on patient wellbeing and health-care costs. Therefore, the progressive nature of COPD and the risks related to an exacerbations ask for continuous patient monitoring. This is confirmed in [4], where the Authors verified a lower rate of exacerbations and related hospitalizations for COPD patients followed with the help of multi-parametric remote monitoring system compared to patients followed using the standard protocol. The continuous monitoring leads, in turn, to the clinical need for an objective way to confirm exacerbation onset. Since biomarkers are not able to reliably differentiate stable COPD variations from exacerbations [5], current attention has focused on symptom scores [6], but symptoms are, by definition, subjective. Therefore, to the best of our knowledge, an effective method for monitoring and detecting the exacerbations during therapy does not exist.

In order to tackle this issue recent research has proposed methods for computer-aided assistance, including event detection, alerting, and monitoring, which can be included in a general platform for chronic disease management at a distance from the hospital [3], [7], [8].

In [3] the Authors observed that exacerbations are known to be associated with changes in the peak-expiratory flow (PEF), but its absolute changes are small and could not reliably detect exacerbations when used in isolation. Therefore, they proposed to differentiate COPD variations from exacerbations through the analysis of changes in heart rate (HR) and oxygen saturation (SpO₂), measurable using a pulse oximeter. They performed a pilot study on 40 COPD patients using a score given by the combination of the two aforementioned physiological parameters which were monitored. The experiments show that changes in heart rate and oxygen saturation differentiate exacerbation onset from stable COPD with sensitivity and specificity ranging in [14%, 100%] and [54%, 94%], respectively. To get these results the Authors analyzed how different values of the threshold on the score affect such performance values. However, the Authors considered the whole population and did not estimate the proposed score performances according to a cross validation scheme, therefore introducing a bias in the results.

In [7], the Authors developed a system for the prediction and detection of exacerbations based on a Bayesian network that requires the measurement of several input parameters. Furthermore, these parameters are both physiological and psychological, such as questionnaires filled out by the patient via a hand-held device. However, most of the people affected by COPD are elderly. For this reason they are not so much collaborative when it is asked to carry out many measures several times a day. Moreover, they may not be accustomed to
high-tech devices and they may experience troubles with touch screen devices. This is due, for instance, to the small fonts shown on the screen of the device or the lack of tactile feedback. In addition, the bayesian network used in this work gave a large weight to the PEF variations, although its variations should be not a discriminating factor for exacerbations [5].

In [8] the Authors proposed a classifier, based on an artificial neural network, which indicates if the forced oscillation technique (FOT) parameter suggests the presence of COPD or not. However, the devices used for FOT measurement are expensive and bulky and this technique is, hence, not suited to be applied at home.

The aforementioned observations show that the identification of an event prefiguring an exacerbation onset could prevent its serious impact on patients’ health status. For the sake of brevity, these events as referred to as worrisome events (WEs) in the following. Hence, in this work we deal with the early identification of WEs to allow the medical doctors to promptly intervene and to manage the situation. To this aim, we develop a computer-aided assistance model able to give more importance to one measure with respect to the other. A-priori knowledge limits the space search for such weights which must have an opposite sign since WEs are associated with HR rise and SpO2 fall. We experimentally investigate if the weighted combination of these signals allows us to successfully identify the WEs on a data set where medical doctors manually labeled the records that could prefigure an exacerbation onset. We perform the tests using a cross validation scheme, which permits us to assess how the results generalize to an independent data set.

II. MATERIALS AND METHODS

A. Dataset

From January 2011 to June 2011 seven patients affected by COPD (4 men and 3 women, aged between 55 and 75 years), and recruited at University Hospital of Università Campus Bio-Medico di Roma were monitored at home via a tele-monitoring application, consisting of a mobile phone and a pulse oximeter (Nonin Medical Inc.). This choice has been motivated by the cohort of elderly patients we used in our study. Indeed, previous experience suggest us that collecting a large number of measures could cause the aversion of the monitored patients which in turn leads to a drop of their collaboration.

The monitoring protocol required that each patient daily measures his/her values of HR and SpO2 with a pulse oximeter connected with the smartphone that sent the data directly to the medical center. Such measurements should begin when the patient wakes up and should be repeated every three hours until the patient goes to sleep. Collected measurements were organized in a data set, where a medical specialist detected the WEs that prefigure a potential exacerbation onset. These situations can be roughly described as a fall of the SpO2 value along with a rise of the HR. In the dataset we found several consecutive days without any measurement: this happens because the patients did not perfectly follow this acquisition protocol for various reasons, which are mostly related to their health status and forgetfulness episodes. Therefore, for each patient we divided the measurements in sets, where a set is composed of consecutive days of recording. A summary of the dataset is reported in TABLE I, which shows the composition of each patient’s sets in terms of average length, number of records and WE number.

B. Oximetry Score

Distinguishing the daily COPD variations from the WEs is important because its prompt detection should have a relevant and positive effect on patient wellbeing. In this respect, and with particular emphasis on exacerbations, recent research interest has focused on the analysis of quantitative and qualitative data collected via tele-monitoring applications [3], [7], [8]. As reported in section II, the age and the technological exposure of our patient cohort permits to collect only quantitative data using a pulse oximeter According to [3], we use the HR and the SpO2 to differentiate WEs from other COPD related events. The values of these physiological quantities are combined by the Weighted Oximetry Score (WOx), defined as follows:

\[ WO_x(t, \omega_{HR}, \omega_{SpO2}) = \frac{\omega_{HR} \cdot V_{HR}(t) + \omega_{SpO2} \cdot V_{SpO2}(t)}{|\omega_{HR}| + |\omega_{SpO2}|} \]

where \( V_{HR}(t) \) and \( V_{SpO2}(t) \) denote the values of HR and SpO2 recordings collected at time \( t \). The symbol \( \overline{\omega} \) means that the argument has been normalized to zero mean and unit variance. Differently from [3], we introduce in the oximetry score definition the two weights \( \omega_{HR} \) and \( \omega_{SpO2} \) in order to give more importance to one measure with respect to the other. A-priori knowledge limits the space search for such weights which must have an opposite sign since WEs are associated with HR rise and SpO2 fall. We therefore assume that \( \omega_{HR} > 0 \) and \( \omega_{SpO2} < 0 \). To further simplify the formula, we can normalize it with respect to \( \omega_{HR} \), obtaining:

\[ WO_x^*(t, \omega) = \frac{\overline{V_{HR}(t)} + \omega \cdot \overline{V_{SpO2}(t)}}{1 + \omega} \]

where the normalized weighted oximetry score \( WO_x^* \) is given by \( WO_x/\omega_{HR} \), and the weight \( \omega \) is given by \( |\omega_{SpO2}|/|\omega_{HR}| \). The computer-aided assistance model detects a WE when the normalized weighted oximetry score \( WO_x^* \) exceeds a certain threshold \( \theta \).

III. EXPERIMENTAL ANALYSIS

A. Protocol

The normalized weighted oximetry score \( WO_x^* \) depends on the value of the weight \( \omega \) and, furthermore, its performances vary also with the alarm threshold \( \theta \). To investigate if \( WO_x^* \) successfully identify the WEs we run the experiments using a cross validation approach, permitting us to assess how the results will generalize to an independent data set. This approach
allows also to overcome the limitations of the experiments presented in [3], which are mentioned in section I.

We consider one patient at time, and we perform a set-out cross validation. Assuming that data related to $i$th patient are composed of $n$ sets, we proceed as described in Algorithm 1, where we run a grid search over the values of $\omega$ and $\theta$ parameters. Their values vary in $[1,20]$ and $[0,10]$, respectively. For $\omega$ the interval has been sampled with unitary step, whereas for $\theta$ the step is equal to 0.5. The interval chosen for $\omega$ permits us to analyze what happens when we give more importance to a fall of SpO$_2$ than to a rise of HR. We do not consider the opposite case, i.e. more importance to a rise of HR than a fall of SpO$_2$, since the a-priori domain knowledge of the medical experts permits us to conclude that this hypothesis does not make sense. It is worth observing that this consideration is valid for the severe COPD affecting our cohort of patients, where exacerbations significantly reduce the SpO$_2$ value. In case of $\theta$ we determine the range interval to be sampled according to the following observation. Reasonable values of HR and SPO$_2$, together with the tested values of $\omega$, entail a maximum value of WO$_2^*$$(t)$ that rarely exceeds the value of 10.

Algorithm 1 Set-out cross validation

- Define the parameter space $\Omega = \{\omega_k\}_{k=1}^m$, where $m$ is the number of tested values.
- Define the parameter space $\Theta = \{\theta_r\}_{r=1}^p$ where $p$ is the number of tested values.

for all the $n$ sets of the $i$th patient do
- Define the test set as the $j$th set, with $j = 1,\ldots,n$.
- Build a reference set merging all sets except the $j$th set.
- For the reference set compute the mean values $\delta_{HR}$ and $\delta_{SpO_2}$ of $V_{HR}(t)$ and $V_{SpO_2}(t)$, respectively.
- For the reference set compute the values of the standard deviation $\sigma_{HR}$ and $\sigma_{SpO_2}$ of $V_{HR}(t)$ and $V_{SpO_2}(t)$, respectively.
- For the test set, normalize the values of $V_{HR}(t)$ and $V_{SpO_2}(t)$ according to the following formulas

$$V_{HR}(t) = \frac{V_{HR}(t) - \delta_{HR}}{\sigma_{HR}} \quad (3)$$
$$V_{SpO_2}(t) = \frac{V_{SpO_2}(t) - \delta_{SpO_2}}{\sigma_{SpO_2}} \quad (4)$$

for all pairs $(\omega_k, \theta_r) \in \Omega \times \Theta$ do
- Compute the value of $WO_2^*(t, \omega_k)$ using $\omega_k$ according to 2.
- if $WO_2^*(t, \omega_k) > \theta_r$ then
  - the sample is labeled as a worrisome event (positive)
else
  - the sample is marked as negative
end if
- Compare the achieved labels with the ground truth and estimate the confusion matrix.
end for
- Sum up the confusion matrices of the $n$ sets in a single confusion matrix.

![Fig. 1. Average sensitivity values achieved for different pairs of the parameters $\omega$ and $\theta$.](image)

![Fig. 2. Average specificity values achieved for different pairs of the parameter $\omega$ and $\theta$.](image)

B. Results

The results have been evaluated using the confusion matrices computed for each patient according to the procedure described in Algorithm 1. Each matrix contains the four values describing the classification outputs of a binary discriminative function. The records labeled as positive ($P$) by the medical doctor, which correspond to WEs, will be named as true positive ($TP$) if the classification system correctly classifies them, whereas they will be named false negative ($FN$) when the classification system incorrectly assigns them to the negative class. Conversely, the records labeled as negative ($N$) by the medical doctor will be named as true negative ($TN$) or false positive ($FP$), whether the classification system assigns them to the negative or positive class, respectively. On these positions, the accuracy of the system is computed from the aforementioned confusion matrix as $\frac{(TP+TN)}{(P+N)}$. However, in our dataset the number of records corresponding to WEs is much smaller than the number of records belonging to the negative class. In these cases where the a-priori sample
distribution is skewed among classes the accuracy is not a reliable performance metric [9]. Thereby, we computed other two metrics describing the performances of the system for each class: they are the sensitivity \((TP/T)\) and the specificity \((TN/N)\).

Averaging out the results achieved among the various sets of the different patients we computed the aforementioned three performances figures. The corresponding results are presented in Fig. 1, 2, 3 and in TABLE II. The three figures show the average values of sensitivity, specificity and accuracy achieved for different values of the pair \((\omega; \theta)\), as described in section III-A. The color bar on the right of each figure defines the mapping of the colors of a figure with the value of sensitivity, specificity and accuracy, respectively.

Fig. 1 and 2 show that a small value of \(\theta\) provides the best performances in terms of sensitivity. Furthermore, the higher the value of \(\theta\), the larger are the performances in terms of specificity. This behavior should be explained observing that a low value of \(\theta\) involves a large number of positive classifications, whereas a high value leads to a large number of negative classifications. Fig. 3 shows that the largest accuracy (91.1\%) was achieved for an intermediate value of \(\theta\) (2) and with a value of \(\omega\) equal to 18. However, due to the high skewness of the considered dataset towards the negative class, this pair of parameter did not permit to achieve good performance in terms of sensitivity, which was equal to 57.1\%. Such a low value is not acceptable in the clinical routine, because it means that about the 43\% of the WE records are not identified. On the other hand, the pair \((0, 8)\) providing the largest sensitivity (80.0\%), caused a drop of the accuracy to 57.6\%. Indeed, such a low value of \(\theta\) entails that many records were erroneously classified as positive.

In order to further support the discussion, we reported in TABLE II the results achieved for each patient. The table shows the best values of sensitivity, specificity and accuracy, where the best value of a metric is reported together with the corresponding values of the other two metrics, achieved with the same values of the two parameters \((\omega; \theta)\). Except for the results of the patients P4 and P5, which show satisfactory performances in terms of all the three metrics, we observe that remarkable joint results of sensitivity and accuracy cannot be found for any other patients.

Finally, we observe that in order to differentiate exacerbation onset from day-to-day symptom variation, in [3] the oximetry score was defined by setting \(\omega = 1\) in equation 2. The results reported in that paper show a large value of sensitivity (100\%) or specificity (94\%) but not a simultaneous large value of such metrics. Our experiments confirm these results, although the goals of our work and the experimental protocol, which is based here on cross validation, differ from [3].

In summary, our results suggest that thresholding the weighted oximetry score cannot adequately differentiate WEs which prefigure an exacerbation onset from other COPD related events. Even though this conclusion may be affected by the limited number of tested patients, we deem that an effective discrimination between WEs and day-to-day variations cannot disregard the temporal evolution of any monitored parameter.

### IV. Conclusions

In this paper we evaluated a method for detecting worrisome events prefiguring an exacerbation onset in COPD patients. We used only data collected via a pulse oximeter, i.e. HR and \(\text{SPO}_2\), thus permitting an easy application of the method to a cohort of patients composed of elderly people affected by severe COPD. In order to achieve a unique measure which takes into account the two monitored parameters, we extended the oximetry score proposed in [3] by adjusting the relative weight of a parameter with respect to the other. We performed a cross validation-based evaluation, which permits us to assess how the results generalize to an independent data set. The reported results suggest that thresholding the

### TABLE II. RESULTS FOR EACH PATIENT. THE BEST VALUE OF A METRIC IS REPORTED TOGETHER WITH THE CORRESPONDING VALUES OF THE OTHER TWO METRICS, ACHIEVED WITH THE SAME PARAMETER VALUES.

<table>
<thead>
<tr>
<th>Patient</th>
<th>Max sensitivity(%)</th>
<th>Specificity(%)</th>
<th>Accuracy(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>88.9</td>
<td>49.0</td>
<td>56.1</td>
</tr>
<tr>
<td>P2</td>
<td>96.3</td>
<td>50.2</td>
<td>61.4</td>
</tr>
<tr>
<td>P3</td>
<td>98.2</td>
<td>50.7</td>
<td>57.5</td>
</tr>
<tr>
<td>P4</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>P5</td>
<td>100</td>
<td>64.6</td>
<td>77.7</td>
</tr>
<tr>
<td>P6</td>
<td>100</td>
<td>50.1</td>
<td>62.5</td>
</tr>
<tr>
<td>P7</td>
<td>87.9</td>
<td>66.0</td>
<td>49.4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Patient</th>
<th>Sensitivity (%)</th>
<th>Max specificity(%)</th>
<th>Accuracy(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>24.1</td>
<td>100</td>
<td>72.4</td>
</tr>
<tr>
<td>P2</td>
<td>22.5</td>
<td>100</td>
<td>80.3</td>
</tr>
<tr>
<td>P3</td>
<td>34.8</td>
<td>100</td>
<td>83.2</td>
</tr>
<tr>
<td>P4</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>P5</td>
<td>99.8</td>
<td>100</td>
<td>99.8</td>
</tr>
<tr>
<td>P6</td>
<td>6.0</td>
<td>100</td>
<td>78.3</td>
</tr>
<tr>
<td>P7</td>
<td>0</td>
<td>100</td>
<td>93.4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Patient</th>
<th>Sensitivity(%)</th>
<th>Specificity(%)</th>
<th>Max accuracy(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>31</td>
<td>100</td>
<td>86.7</td>
</tr>
<tr>
<td>P2</td>
<td>74.5</td>
<td>98.3</td>
<td>92.9</td>
</tr>
<tr>
<td>P3</td>
<td>46.8</td>
<td>96.4</td>
<td>80.3</td>
</tr>
<tr>
<td>P4</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>P5</td>
<td>99.8</td>
<td>100</td>
<td>99.8</td>
</tr>
<tr>
<td>P6</td>
<td>78.4</td>
<td>97.2</td>
<td>93.1</td>
</tr>
<tr>
<td>P7</td>
<td>0</td>
<td>100</td>
<td>93.4</td>
</tr>
</tbody>
</table>
weighted oximetry score cannot effectively differentiate the worrisome events from other COPD related events. Indeed, since joint satisfactory values of sensitivity and accuracy cannot be achieved by the evaluated method, we concluded that it is not adequate to perform a pre-selection of the cases that have to be examined, which would enable the physician to focus his/her attention only on relevant cases.

In order to overcome the observed limitations, future works are directed towards the development of a model which analyzes the temporal evolution of the monitored parameters. We deem that this will permit us to perform a deeper analysis of the disease’s course, thus improving WE detection.

ACKNOWLEDGEMENTS

This work has been carried out within the SweetAge project, which has been funded by the Lazio Region through the “Finanziaria Lazio per lo Sviluppo (FILAS)”.

REFERENCES


