Topic models-based features from nighttime data for classification of patients with COPD

Gabriele Spina, Pierluigi Casale, xxx, xxx, and Albertus C. den Brinker

***Abstract*— Night-time symptoms are important indicators of impairment for many diseases and particularly for respiratory diseases as Chronic Obstructive Pulmonary Disease (COPD). This work introduces a technique for predicting the pathological condition in patients with COPD using features extracted from multimodal sensor data during night-time only. Sensor data were discretized and presented as the units composing symbols that describe subjects' night-time. The co-occurrence of these symbols in different ways and proportions during the night creates thematically coherent groups describing particular sleep modalities used for classification. We demonstrated the capabilities of our approach by applying it to a real-world COPD patient cohort showing its validity in assessing sleep in relation to the pathological condition. The results showed that it is possible to differentiate between healthy and patients with COPD with 94% accuracy and between disease severity and dyspnoea severity with an accuracy of 94% and 93%, respectively.**

***Index Terms*—COPD, Topic models, classification, sleep.**

I. Introduction

CHRONIC Obstructive Pulmonary Disease (COPD) is currently the third leading cause of death worldwide [1]. It is caused, among others, by smoking and air pollution and it is characterized by chronic inflammation of the lung airways, and degradation of lung tissue which result in airflow limitation [2]. COPD is a global health problem because of its high prevalence, increasing incidence, and associated socio-economic costs [3].

However, COPD is severely underdiagnosed and therefore undertreated [4]. Spirometry is compulsory for the diagnosis of COPD [3]. Spirometry is now cheap and widely available, but time availability and quality control are often said to limit its implementation, particularly in primary care where there is a great deal of controversy regarding the quality of the tests performed by non-expert professionals [5]. Although detecting the disease at an early stage can increase the survival rate [6], obtaining spirometry for each smoker with or without symptoms of dyspnoea, cough, or sputum production is not a feasible solution. With the advent of health-sensor technologies and advanced data analysis methods there is a gradual permeation of these technologies into actual health care and patients' homes enabling new care services and their use as supportive systems [7].

In this work we show that the use of multimodal sensor modalities acquired during night-time only provides a good predictor of the presence of the disease in patients with COPD. In particular, we 1) used data coming from a single activity monitor worn on the upper arm to define new sleep modalities using a data driven methodology, and 2) we showed that these sleep modalities are valid constructs to assess sleep of patients with COPD in relation to their pathological condition. The defined sleep modalities were able to differentiate between nights of patients with COPD vs. nights of healthy subjects and, more in detail, predict the level of the disease and dyspnoea severity. It is worth mentioning that, for the best of our knowledge, to date, this is the first study that introduces the use of night-time data only derived from a single activity monitor for diagnosis of COPD. The objective is to reduce the uncertainty of predictions by fusing multimodal information and the use of data mining techniques.

II. Related Works

Pervasive healthcare may enable a paradigm shift from the established centralized healthcare model to a pervasive, user-centred and preventive health management [8]. Wearable and unobtrusive technologies make more health-related data available than ever before enabling caregivers to make decisions on a broader basis of information. In clinical settings, data analytics and decision support systems have been established in several data-rich environments [7], among which COPD management.

A. Automatic classification of patients with COPD

Different machine learning techniques, such as artificial neural networks [9], and multiple instance learning [10] have been used for automatic recognition of COPD based on high resolution computed tomography scanning, which enables the direct evaluation of the lungs and airways. In [11] the authors proposed a method to predict the condition of a patient with COPD at an early stage and in absence of clinician. The predicted physiological parameters such as forced expiratory volume in the first second expressed as percentage of a

This paragraph of the first footnote will contain the date on which you submitted your paper for review. It will also contain support information, including sponsor and financial support acknowledgment. For example, "This work was supported in part by the U.S. Department of Commerce under Grant BS123456".

The next few paragraphs should contain the authors' current affiliations, including current address and e-mail. For example, F. A. Author is with the National Institute of Standards and Technology, Boulder, CO 80305 USA (e-mail: author@ boulder.nist.gov).

S. B. Author, Jr., was with Rice University, Houston, TX 77005 USA. He is now with the Department of Physics, Colorado State University, Fort Collins, CO 80523 USA (e-mail: author@lamar.colostate.edu).

T. C. Author is with the Electrical Engineering Department, University of Colorado, Boulder, CO 80309 USA, on leave from the National Research Institute for Metals, Tsukuba, Japan (e-mail: author@nrim.go.jp).

maximum value (FEV1%), six-minute walking test, Modified Medical Research Council (MMRC) dyspnoea scale, and body mass index (BMI) are weighed using particle swarm optimization and the state of the patient is fuzzified in accordance to the GOLD criteria. Although useful, all these methods require signals which are only available in hospitals or laboratories and can currently not be provided by automated wireless remote detection.

Newandee et al. [12] have studied COPD severity classification using principal component and cluster analysis on heart rate variability (HRV) parameters using heart rate, blood pressure, and respiration signals. Results demonstrated that these two groups could be differentiated with greater than 99.0% accuracy. Furthermore, differences on the same HRV parameters between all four severity levels of COPD subjects were also investigated. These groups were differentiated with over 88.0% accuracy. However, data were acquired under controlled laboratory conditions.

The CHRONIOUS system [13] offers an intelligent system for the analysis and the real-time evaluation of patient's condition. A hybrid classifier has been implemented on a personal digital assistant, combining a support vector machine, a random forest, and a rule-based system to provide a categorization scheme for the early and in real-time characterization of a COPD episode. This is followed by a severity estimation algorithm which classifies the identified pathological situation in different levels. The achieved characterization accuracy has been found 94%. Sensor data were acquired by external sensors and several devices attached to a wearable jacket during daytime making the system rather cumbersome and difficult to be used in clinical practice.

In our previous work [14] it has been shown that probabilistic daytime activity biomarkers derived from an unobtrusive activity monitor data by using topic modelling techniques are able to cluster subjects with and without COPD with 86% of accuracy. Patients' monitoring and classification using night-time data may offer several advantages since, spending one third of their life with sleeping, it belongs to one of the prime activities humans pursue in which disease trends can be better observed. Moreover, compared to daytime hours, sleeping hours may offer a better trade-off between patients' comfort, sensor unobtrusiveness and signal quality [8].

B. Sleep in patients with COPD

Sleep disturbance, such as sleep fragmentation during the night, is common in patients with COPD [15], and is a major complaint after dyspnoea and fatigue [16]. Despite the high prevalence of disturbed sleep in COPD, night-time symptoms are often underestimated and are not a focus of current disease management [15].

Sleep in patients with COPD is usually assessed using a patient-completed diary that consists of asking patients to record their sleep duration, recalled sleep disruptions, and a sleep score that reflects the degree of perceived sleep disruption [17]. Self-reported measures of sleep duration and quality provide a useful insight into the patient's perception of the nocturnal burden of their disease, but their precision and reliability are poor compared to objective measures of sleep [15].

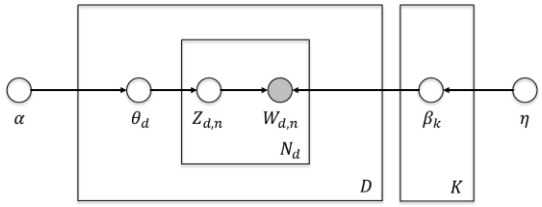

Fig. 1 Graphical model for LDA. Each node is a random variable, edges denote possible dependences. The only observed variables (shaded) are the symbols (W). The distribution of the symbols in a sleeping modality (β) and the distribution of the sleeping modalities during a night (θ) depend only on the sleeping modalities hyperparameters η and α that control the mean shape and sparsity of the distributions. Z represents the symbol sleeping modality assignment.

Sleep laboratory measurement tests, such as polysomnography (PSG), offer a well-validated, reliable and reproducible method of collecting sleep data, but they are expensive, intrusive and require patients to attend a clinic for overnight recording.

The use of wearable sensors has been mainly limited to coarse-grained methods such as actigraphy, for which limb motions are logged providing some insight in patient's sleep. Actigraphy correlates well with PSG in differentiating sleep from wake [18], but rather than a replacement for PSG, it should be regarded as another means for assessing sleep, particularly when sleep architecture and extensive physiological monitoring are not necessary.

Activity monitors are useful tools that are becoming popular to objectively assess the sleep–wake cycle. They provide minimally invasive measures of the continuity and hence quality of sleep and other physiological measurements such as energy expenditure, body temperature and galvanic skin response. In this work, metabolic and physiological data recorded during night-time using an activity monitor are symbolized and presented as the "letters" composing the "words" that describe the night of a subject. The co-occurrence of these words in different ways and proportions during the night creates groups of words describing the modalities in which a subject sleeps. Using these data we extracted patients' sleep modalities that were valid to assess sleep in relation to the presence of the pathological condition. While previous studies aimed at finding differences between healthy individuals and patients with COPD [19], this study seeks to evaluate severity classification of the disease. This is a challenge because differences in objective sleep measures between COPD classes are subtle [19, 20] making them difficult to detect. The proposed methodology, exclusively defined using data coming from one unobtrusive device, is able to differentiate between COPD and healthy-type of nights, and to discriminate between different GOLD grades and MMRC scores.

III. BACKGROUND

Topic models are algorithms for discovering the hidden grouping variables that pervade a large and unstructured collection of documents. LDA is an example of a topic model [21] in which data are treated as observations arising from a generative probabilistic process. In the context of text modelling, given a set of topics defined as distributions over

words, the generative process populates the documents with words such that the documents have a particular desired thematic structure. Beside its generative process, LDA can also be used to calculate the hidden variables that likely generated the collection of documents. One of the ways to achieve this is to use variational inference to approximate the posterior distribution over the hidden variables defined by LDA. In a nutshell, variational inference posits a parametrized family of distributions over the hidden structure, and then, finds the member of that family that is closest to the posterior according to the *Kullback–Leibler* divergence. The intuition behind using LDA [21] for sleep monitoring is that each night is a mixture of thematically coherent measures just as a text document is a mixture of thematically coherent words. The graphical model for LDA is shown in Fig. 1. All the assessed nights ($d_{1:D}$) share the same set of sleep modalities ($\beta_{1:K}$) that are defined as Dirichlet distributions over the observed set of symbols (W) which are the terms of a fixed vocabulary. The observed symbols (input of the model) are composed by multimodal measures coming from the sensors of an activity monitor. Each assessed night exhibits sleep modalities in different proportion providing an explicit finger print θ. In particular, each night is a different distribution ($\theta_{1:D}$) over the sleep modalities activation probabilities that also follows a Dirichlet distribution. In such a model, the N symbols ($W_{d,n_{1:N}}$) that compose the D nights are the only random variables observed and depend on the per word sleep modality assignment ($Z_{d,n}$) and all the β_k . Each sleep modality then is composed indirectly by low-level sensor measures that belong, with a certain probability distribution, to different thematic areas. Our hypothesis is that nights related to different group of subjects would have a different distribution over the sleep modalities that in turn would be composed by different distributions over symbolic words defined combining discretized sensor measurements. The proposed methodology will first extract all the β_k in a data driven fashion using data from a subset of COPD patients and healthy subjects and then, for each assessed night, it will calculate a probabilistic feature vector θ composed by the histogram of activation probabilities of all the sleep modalities $\beta_{1:K}$. These probabilistic features will be used to classify a large cohort of patients.

IV. Methods

In the application of LDA, a word, defined to be an instance of a vocabulary, is considered as the basic unit of discrete data. In this work, multimodal monitoring signals such us metabolic equivalent of task (MET) [22], temperature, galvanic skin response and number of steps were first converted into a discrete alphabet of letters and then combined into symbolic words. Our approach in selecting the letters and then the words composing the vocabulary benefits from a methodology that preserves the interpretability of the vocabulary and that allows the generation of symbols that actually do not occur in the current documents. The multidimensional data space is first divided in subspaces according to the METs values in order to conveniently define a repertoire of physical activities in which a person may participate [22]. In each subspace, we divided temperature and galvanic skin response data into partitions (letters) that are then combined to form string of symbolic words. We extract the vocabulary in a subset of COPD and healthy patients and show that the constructed vocabulary is able to model the data of a much larger cohort of patients. The dataset and the methodology developed are described in detail in the following sections.

A. Participants

Data from 1384 patients from ten countries (United Kingdom, Ireland, The Netherlands, Germany, Switzerland, Italy, Spain, The United States of America, Brazil, and Australia) diagnosed with mild to very severe COPD were pooled from previous studies (references are omitted for the sake of brevity) and considered for analysis. Participants were included if they had COPD with a post-bronchodilator ratio of forced expiratory volume in the first second (FEV_1) to forced vital capacity (FVC) < 0.70 and they were clinically stable (i.e., stable shortness of breath and sputum production). We report baseline data recorded before any specific interventions were undertaken. Centers from The Netherlands and UK also provided data on 66 healthy control subjects that were matched for age, gender, and BMI with a subgroup of 66 COPD patients. On the basis of a 1:1 multivariate matching, the closest possible *case:control* matches were determined. Subjects matched exactly for age and gender, the median error between BMI values of matching subjects was 0.58 [0.29–1.2] Kg/m^2. Subject group characteristics are presented in Table 1. The data collection was approved by ethics committees at each of the participating centers, according to local regulations. Written informed consent was provided by all participants.

B. Data recordings

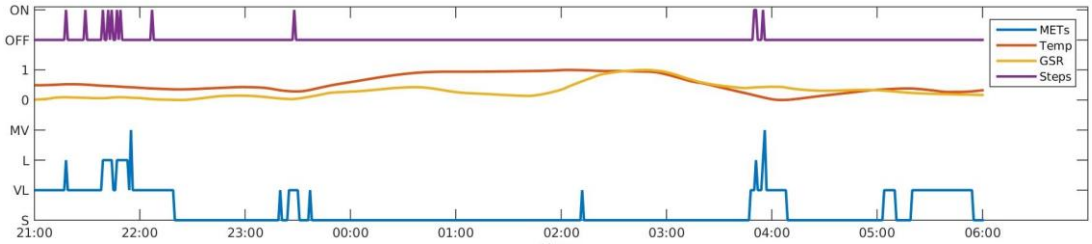

Fig. 2 Nighttime data from 21:00 pm to 06:00 am of a participant in the study. In blue Metabolic Equivalent of Task (MET), in yellow Galvanic Skin Response (GSR), in red Temperature (Temp), in purple Steps. METs data, defined as the energy cost of physical activities as a multiple of the resting metabolic rate, are divided in 4 intensity category (S, VL, L, and MV). GSR and temperature data are scaled in the interval 0-1 for visualization. Steps are discretized in a binary form (ON = at least 1 step performed, OFF = no steps performed).

TABLE 1
SUBJECT GROUP CHARACTERISTICS

	All COPD*	Matching healthy	Matching COPD**
	n = 1059	n = 66	n = 66
Male/Female (n)	689/370	30/36	30/36
Age (years)	66 [61–72]	65 [61–70]	65 [61–70]
BMI (Kg/m2)	25.9 [22.5–29.6]	25.2 [23–27.3]	25 [22.5–27.8]
FEV1 (% predicted)	49 [34–64]	107 [97–117]	42 [29–63]
GOLD 1–2–3–4 (n)	93–419–354–193	-	8–16–23–19
MMRC 0-1-2-3-4 (n)	145-279-228-195-67	44-1-0-1-0	6-15-18-15-5
Assessed nights (n)	6446	404	411
Weekdays (%)	67.3	69	65
Nights per subject	6 [6–6]	6 [6–6]	6 [6–6]

Data are summarized as absolute frequency (n), relative frequency (%), or median and quartiles [Q1–Q3]. *MMRC data for 914 subjects. **MMRC data for 59 subjects.

Study participants wore the SenseWear Armband or SenseWear Mini Armband activity monitors [23] (BodyMedia Inc., Pittsburgh, PA, USA) on the upper arm both during daytime and night-time so that continuous, real-life data were recorded in a natural environment. These devices included an accelerometer with different physiological sensors: a heat flux sensor, a galvanic skin response (GSR) sensor, a skin temperature (ST) sensor, and a near-body ambient temperature sensor [24, 25]. Data were sampled at one minute intervals and, together with demographic characteristics, were used to estimate METs using proprietary algorithms developed by the manufacturer. The use of multisensory data in combination with pattern recognition algorithms ensured that the MET estimation was insensitive to noise and random motion artefacts [26]. For each minute, the device also recorded steps count, information about the sleeping status of a patient (0=awake, 1=sleeping), and posture (0=lying down, 1=not lying down). Night-time sleep was defined as sleep that occurs between 21:00 and 06:00 [27]. For this analysis we considered only MET, ST, GSR, SC, and Sleeping Status (SL) data within this time interval. The following night-time sleep measures were also derived from the sleep status information provided by the sensor: night sleeping time, number of nocturnal sleeping bouts and duration of nocturnal sleeping bouts. Sleeping bouts were defined as consecutive minutes marked by the sensor as sleeping. Participants who wore the device for at least four nights (two weekdays + Saturday + Sunday) were included [25]. In total 1059 patients with COPD and 66 healthy controls were included in the analysis. The median number of nights analysed per patient was six (four during weekdays, two during weekends), resulting in a total of 6446 valid nights assessed, of which 4335 (67.3%) were during weekdays and 2111 (32.7%) during weekends.

C. Topic Models

Sensor data from the 66 healthy subjects and the matched 66 COPD patients subsample were used to create the vocabulary of words. METs data were divided into intensity categories (*IC*) using the thresholds proposed by the American College of Sports Medicine: very light intensity (*VL*), < 2.0 METs; light intensity (*L*), 2.0 to 2.9 METs; and moderate-to-vigorous intensity (*MV*), ≥ 3.0 METs [12]. Minutes marked by the sensor as sleeping and with METs < 2.0 formed a separated category named sleeping (*S*). Step counts were converted in a binary form depending on whether the participant performed steps in each assessed minute (0=no steps performed, 1=steps performed). Temperature and galvanic skin response data were first cleaned from missing values and outliers (i.e. temperature values outside the range [24-40 °C] and GSR values outside the range [0-8 µSiemens]) and then, for each subject, were centred across the mean over multiple assessed nights. An example of data stream for a single patient night can be found in Fig. 2.

In order to have sparse sleep modalities and symbols that best represent the original signal, it is desirable to have a discretization technique that produces symbols with equal probabilities [28] and that minimizes the distortion of the partitioned signal [29]. Therefore, the empirical cumulative distribution functions (ECDFs) of temperature and galvanic skin response data were estimated separately for each intensity category, and the three breakpoints ($Be_{i,i=1:3}$) that divided the data into four equiprobable partitions ($Pe_{j,j=1:4}$) were derived. Separately the breakpoints ($Bd_{i,i=1:3}$) which divided the same set of data in four partitions ($Pd_{j,j=1:4}$) minimizing the mean square distortion of the quantization [29] were also calculated. The final partition breakpoints ($Bf_{i,i=1:3}$) were calculated averaging the corresponding pairs of breakpoints $Bf_i = \frac{Be_i + Bd_i}{2}$ and used to divide the data into four contiguous, non-overlapping ranges of values. Final partition ranges ($Pf_{j,j=1:4}$) are sorted in ascending order such that the first

range (Pf_1) represents partition of data with the smallest values, and the last range (Pf_4) represents data with the highest values. The vocabulary of terms was built by allowing all the possible combinations between partitions ranges of temperature, galvanic skin response data and binary values of steps that share the same *IC*.

For the sleeping category, for example, the 32 terms of the vocabulary describing the sleeping intensity category can be represented by:

$$\begin{aligned} t_1^S&: [S, \quad Pf_1^{ST}, \quad Pf_1^{GSR}, \quad Steps_{No}] \\ t_2^S&: [S, \quad Pf_1^{ST}, \quad Pf_1^{GSR}, \quad Steps_{Yes}] \\ &\vdots \\ t_{32}^S&: [S, \quad Pf_4^{ST}, \quad Pf_4^{GSR}, \quad Steps_{No}] \end{aligned}$$

The sum of terms across different activity levels is the total number of words. In particular, a total of 128 terms were initially created. As a last step, in order to remove nonsense words and increase the frequency of *VL* and *MV* words to obtain sparse topics, we pruned the vocabulary adding wildcard characters [21]. We used a wildcard character to replace the symbol related to the steps performed during sleeping *IC* (i.e. 16 words removed). Considering the neutral wildcard symbols the original 32 terms representing the sleeping category can be represented by 16 terms as:

$$\begin{aligned} t_1^S&: [S, \ Pf_1^{ST}, \ Pf_1^{GSR}] \\ t_2^S&: [S, \quad Pf_1^{ST}, \quad Pf_2^{GSR}] \\ &\vdots \\ t_{16}^S&: [S, \quad Pf_4^{ST}, \quad Pf_4^{GSR}] \end{aligned}$$

Two wildcard characters replaced the temperature and galvanic skin response symbols during *VL* and *MV* intensities (i.e. 60 words removed) to know if the subject was in these two *IC* because of moving. In view of sparsity we also weighted the informativeness of the remaining words in the vocabulary based on their inverse document frequency (IDF) score. Those terms that have a high IDF are considered more informative, because they rarely occur in the collection. In particular, we removed the words occurring in at least 90% of the documents since, occurring so frequently, they are more likely to obscure than facilitate a meaningful decomposition of the collection of documents. The term frequency (TF), usually used in combination with IDF to form the TF-IDF score [30], was not considered since it penalizes words that rarely appear within a document such as words of light or moderate to vigorous intensity, these words are important in the identification of sleep modalities correlated with the disease. The removed term is [L, $Steps_{Yes}$]. The IDF score of the words, and, subsequently, the set of removed words, are related to the wildcarding procedure previously described. If more letters are used, the words created will be more specific with the consequence of a higher IDF score average for the words in the vocabulary. On the other hand, a higher threshold on the IDF score (i.e., IDF equal to the one of the words present in 70% of documents) could cause the removal of all the terms useful to identify specific patients' subtypes. We set a threshold on the IDF equal to the one of the words present in 90% of documents as [14]. The variables for each *IC* and the final associated symbols are shown in Table 2.

For topic discovery, we used the LDA implementation of [21], and we considered each day of assessment as a separate document. Pre-processed data were received by a symbolization unit which maps the raw, multivariate, continuous-time data stream into a signal which can be handled by LDA. In particular it maps the data contained into each 4-elements vector $V = [IC, ST, GSR, SC]$ representing one assessed minute into a set of discrete symbols which could be interpreted as the letters of text words. Each assessed minute was mapped with an instance of the vocabulary by associating the selected values in V with their partitions and then concatenating them. Once that a term of the vocabulary was assigned to each minute, documents were created by constructing for each day a histogram of terms. We computed the results from a number of sleep modalities that varies from 3 to 20, and set the hyperparameter α equal to 0.01 as in [31]. Hyperparameters are optimized with a variational expectation maximization algorithm initialized by randomly choosing a small number of "seed" documents [30]. We selected 18 seeds (nine from healthy subjects and nine from COPD patients). Routines did not change in their overall composition with different seed sets. Once the routines were calculated and each assessed minute of each night mapped to a symbol of the vocabulary we inferred each night in order to estimate the minutes spent in each routine.

TABLE 2
VARIABLES SELECTED FOR EACH INTENSITY AND ASSOCIATED SYMBOLS

Sleeping			Very light		
ST	*GSR*	*SC*	*ST*	*GSR*	*SC*
$PF_1\ PF_2\ PF_3\ PF_4$	$PF_1\ PF_2\ PF_3\ PF_4$	*	$PF_1\ PF_2\ PF_3\ PF_4$	$PF_1\ PF_2\ PF_3\ PF_4$	*No Yes*
Light			Moderate to vigorous		
ST	*GSR*	*SC*	*ST*	*GSR*	*SC*
*	*	*No Yes*	*	*	*No Yes*

V. CLASSIFICATION

The evaluation of the results for classifying subjects based on their pathological condition (66 healthy vs. 66 COPD), for distinguishing among different stages of the diseases (healthy, GOLD 1, GOLD 2, GOLD 3, GOLD 4) and different dyspnoea scores (MMRC 0, MMRC 1, MMRC 2, MMRC 3, MMRC 4) is performed through a 10-fold random-partitioning cross-validation process.

We firstly divided the data from the 132 subjects, comprising healthy subjects and matched COPD patients, into ten subsets. At each iteration, one of the subsets was used as the test set and the other nine subsets formed a training set. From each subject in the training set we randomly selected one night, represented by its characteristic vector of activation probabilities $\theta = [\theta_1, \ldots, \theta_k]$ over the sleep modality $\beta_{1:K}$. We used these distributions to compute a square dissimilarity matrix A between pairs of nights according to *Kullback–Leibler* divergence as in [32], in which A(i, j) denotes the

dissimilarity between the i^{th} and j^{th} randomly selected nights $A(i,j) = \sum_k \theta_k^i \log \frac{\theta_k^i}{\theta_k^j}$ with θ_k^i and θ_k^j the activation probabilities of the sleep modality β_k for the nights represented by θ^i and θ^j. The choice of the dissimilarity measure is critical and must fit the nature of the features in question, which in this case are discrete probability functions. Secondly, we calculated the eigenvectors and eigenvalues of A so that $AV = DV$, where D is a diagonal matrix of eigenvalues and V a matrix whose columns are the corresponding right eigenvectors and there are as many eigenvectors and eigenvalues as there are rows in the initial matrix. Eigenvalues were ranked from the greatest to the least. Using the transformation $V_T D_T^{-1}$, with V_T the truncated eigenvector matrix of the first *n* eigenvectors of V and D_T the associated and truncated eigenvalue matrix, we summarized and attempted to represent inter-nights dissimilarities in a lower dimensional space.

We iteratively projected into $V_T D_T^{-1}$ all the nights of each subject in the training set and test set such that the between-object dissimilarities are preserved as well as possible. In particular, given a vector of sleep modality activation probabilities θ representing one night, we calculated the vector $\boldsymbol{x}$ of pairwise dissimilarities between θ and the nights used to compute ${V_T D_T}^{-1}$. Then we assigned to θ a location in a low-dimensional space projecting $\boldsymbol{x}$ into the learned space ${V_T D_T}^{-1}$ according to $\boldsymbol{v}' = \boldsymbol{x} {V_T D_T}^{-1}$. Iterating this operation for each night in the training set and test, the positions of points relative to each other did not change but the coordinate systems changed resulting in a rotation of the data. In a nutshell we created a transformed feature set which rows represent a night of a patient and columns the projection of the pairwise dissimilarities into the space of the first *n* eigenvectors and eigenvalues learned using one single night per subject. We performed class recognition by using a Random Forest (RF) [33] classifier with 50 trees. RFs are ensembles of weakly correlated decision trees that vote on the classification and have been shown to provide good generalization compared to individual decision trees. We used the transformed features representing the nights in the training set to train the ensemble of 50 classification trees. We report the results for the classification of the single nights and of the subjects as results of the vote of his assessed nights.

We applied the same procedure to evaluate the classification of the five disease severity grades (healthy, GOLD 1, GOLD 2, GOLD 3, and GOLD 4) and of the five dyspnoea scores (MMRC 0, MMRC 1, MMRC 2, MMRC 3, and MMRC 4) for a total of 1125 subjects and 690 subjects, respectively. We explored all possible combinations for different number of topics β_k (with *k* varying from 2 to 20) and selecting the first *n* eigenvectors (with *n* varying from 1 to 20).

For comparison we also evaluated the classification performances in the case standard features extracted during night-time such us total night sleeping time, number of nocturnal sleeping bouts and duration of sleeping bouts are used.

VI. Results

A. Healthy vs COPD

The average accuracy across all ten cross validation trials in classifying each night as healthy or COPD-type is shown in Fig. 3.a. A total of 815 nights were classified from the 132 matched subjects. The mean accuracy for the single night classification was 0.89 (SD = 0.008). The maximum accuracy (0.91, SD = 0.04) was achieved setting the number of sleep modalities to 13 and using the first 13 eigenvectors. For this setting we achieved an accuracy of 0.94 (SD = 0.05) for the subject classification.

To compare the time spent in each of the 13 sleep modalities in the 1059 COPD patients and the subgroup used to extract the sleep modalities, we constructed a linear mixed-effect model (LMM) for each sleep modality, with GOLD and MMRC as ordinal explanatory variables; subset group (i.e. all COPD vs matched COPD), smoking status, country of origin, gender and day group (i.e. weekday vs. weekend day) as categorical explanatory variables; age and BMI as continuous explanatory variables. Least Squares means (LS-means) and differences of LS-means of the fixed effects were used to compare the two subset groups. To account for repeated measurements, we used random effects on two levels. On the highest level, we included a random intercept per patient. The second level, within patients, had a random intercept for each day group (weekdays vs. weekends). The residuals then accounted for the differences between days within the same day group.

The model accounts for by-subject and by-day group variability. Degrees of freedom and p-values for significant differences (significant if $p<0.05$) were computed using Satterthwaite's approximation [34]. To construct the models we used the lmer function of the package lme4 in R [35].

Comparison between the time spent in each of the 13 sleep modalities in the 1059 COPD patients and the subgroup used to extract the sleep modalities (Fig. 3.b) shows that there are no statistical differences ($p>0.1$) in the time spent in each sleep modality between the two groups. This indicates that the sleep modalities, created using a subset of patients, are able to generalize across many COPD patients.

B. Disease severity

The average accuracy across all ten cross validation trials in classifying each night as belonging to one of the four disease classes is shown in Fig. 3.c. A total of 6446 nights were classified for a total of 1059 subjects. The mean accuracy for the single night classification was 0.85 (SD = 0.01). The maximum accuracy (0.87, SD = 0.02) was achieved setting the number of sleep modalities to 16 and using the first 9 eigenvectors. For this setting we achieved an accuracy of 0.94 (SD = 0.03) for the subject classification.

C. MMRC score

The average accuracy across all ten cross validation trials in classifying each night as belonging to one of the five MMRC classes is shown in Fig. 3.d. A total of 5588 nights were classified for a total of 914 subjects. The mean accuracy for the single night classification was 0.82 (SD = 0.01). The maximum accuracy (0.84, SD = 0.03) was achieved setting the number of

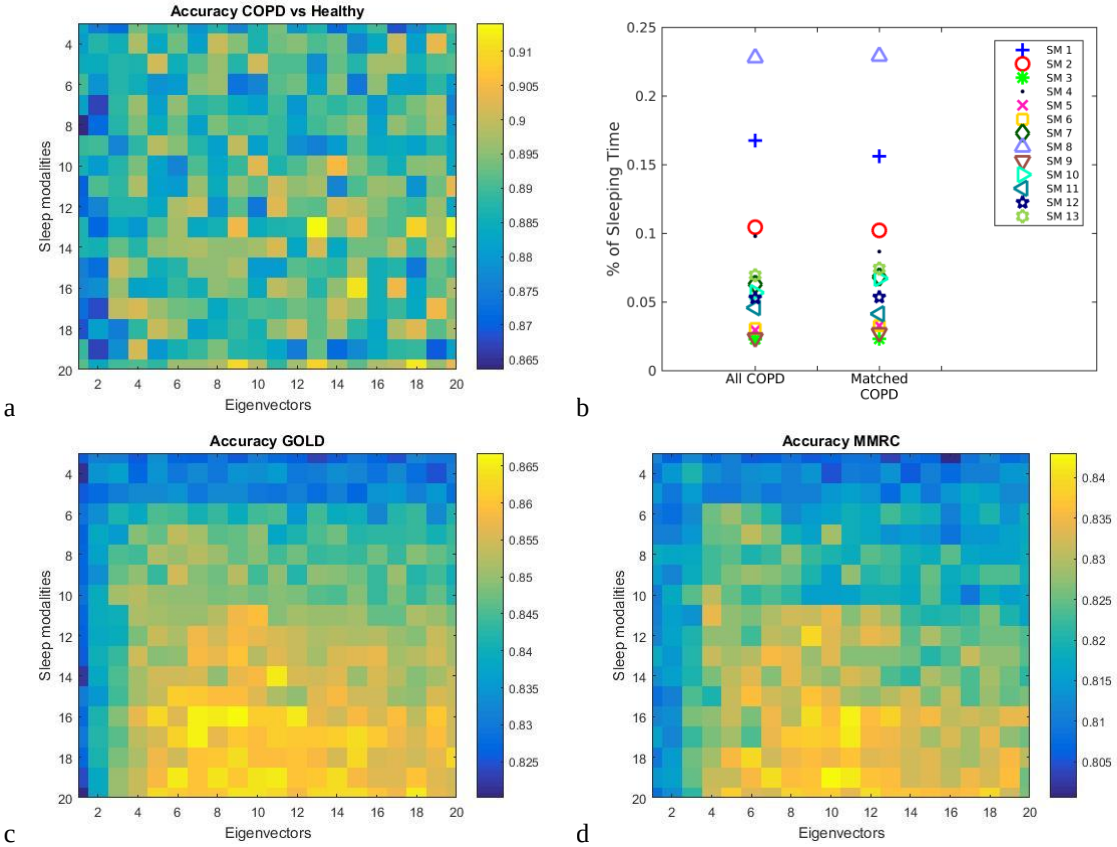

Fig. 3 a) Accuracy matrix for differentiating between nights of patients with COPD vs. nights of healthy subjects varying the number of sleep modalities and the number of eigenvectors. b) LS-means for time spent (as percentage of the night-time assessed) in each sleep modality (SM) for all the COPD patients (n = 1059), and the COPD subset used to generate the sleep modalities (n = 66). c) Accuracy matrix for predicting the patients' level of disease severity given each assessed night. d) Accuracy matrix for predicting the patients' level of the dyspnoea given each assessed night.

sleep modalities to 16 and using the first 9 eigenvectors. For this setting we achieved an accuracy of 0.93 (SD = 0.03) for the subject classification.

D. Classification using standard features

The mean accuracy across all ten cross validation trials in classifying each night as healthy or COPD-type is 0.85 (SD = 0.06). For the subject classification the accuracy was 0.92 (SD = 0.07). The average accuracy across all ten cross validation trials in classifying each night as belonging to one of the four disease classes was 0.58 ($SD = 0.02$). For the subject classification the accuracy was 0.72 ($SD = 0.05$). The average accuracy across all ten cross validation trials in classifying each night as belonging to one of the five dyspnoea classes was 0.57 ($SD = 0.02$). For the subject classification the accuracy was 0.76 ($SD = 0.05$).

VII. Discussion

Early diagnosis of a disease is probably the most valuable asset in order to prevent damages or stall the progression of the disease by effective interventions. Early stages are difficult to assess. More evidence could typically be obtained by longer observations during daily life and obtained with the smallest burden for the patients. For these reasons such observations should ideally be done by unobtrusive means preferably in a familiar environment. Although it is already possible to collect large amounts of data from a single person during a continuous period of his normal life, it is difficult to merge the data into a set of features enabling diagnosis or assistance in diagnosis. We demonstrated the usefulness of our approach by applying it to a real-world COPD patient cohort of more than 1000 patients and a subset of healthy controls showing its validity in assessing sleep in relation to the pathological condition. In particular, we have shown that by fusing multimodal information derived from a device worn exclusively during the night it is possible to differentiate normal subjects from subjects with COPD with an average accuracy of 94% at the subject level (i.e. using several assessment nights to classify each subject) and an average accuracy of 89% in classifying each single night. In addition, patients were classified as well according to their disease severity level and dyspnoea grade with 94% and 93% accuracies, respectively. Standard features such as total night sleeping time, number of nocturnal sleeping bouts and duration of sleeping bouts were able to differentiate between healthy subjects and patients with COPD with good accuracy. However, in agreement with Hartman et al. [20] who did not find significant associations between night's rest parameters and GOLD or MMRC, these features were not able to discriminate between different disease severity stages and dyspnoea grades. Discovered latent structures in night-time data, instead, were sufficiently sensitive to pick up subtle differences existing between the four groups of COPD subjects and five groups of dyspnoeic patients. Based on the target outcome, the settings in

the latent model should be adapted. In particular, a lower number of latent structures is required to get the best classification performance for a two class problem compared to the classification of more classes. For the four classes and five classes problems, a higher number of latent structures, in turn more specific, led to better classification results. It is worth noting that the classification accuracies were always greater than 80% regardless the number of sleep modalities and eigenvectors selected. We believe our contribution represents a step forward towards a better support to the diagnosis of a complicated disease that will hopefully lead to better patient care. With the aim of learning different disease subspaces in mind, an open question for follow-up work is whether it is possible to use (fully or partially) structures provided by clinicians using known clinical relevant features instead of hidden structures extracted from the data.

REFERENCES

[1] R. Lozano, M. Naghavi, K. Foreman, S. Lim, K. Shibuya, V. Aboyans, J. Abraham, T. Adair, R. Aggarwal, S. Y. Ahn *et al.*, "Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the global burden of disease study 2010," *The Lancet*, vol. 380, no. 9859, pp. 2095–2128, 2013.

[2] P. M. Calverley, "Copd: early detection and intervention," *CHEST Journal*, vol. 117, no. 5_suppl_2, pp. 365S–371S, 2000.

[3] J. Vestbo, S. S. Hurd, A. G. Agust, P. W. Jones, C. Vogelmeier, A. Anzueto, P. J. Barnes, L. M. Fabbri, F. J. Martinez, M. Nishimura *ét al.*, "Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: Gold executive summary," *American journal of respiratory and critical care medicine*, vol. 187, no. 4, pp. 347–365, 2013.

[4] M. Bednarek, J. Maciejewski, M. Wozniak, P. Kuca, and J. Zielinski, "Prevalence, severity and underdiagnosis of copd in the primary care setting," *Thorax*, vol. 63, no. 5, pp. 402–407, 2008.

[5] E. Derom, C. Van Weel, G. Liistro, J. Buffels, T. Schermer, E. Lammers, E. Wouters, and M. Decramer, "Primary care spirometry*," *European Respiratory Journal*, vol. 31, no. 1, pp. 197–203, 2008.

[6] R. A. Pauwels, A. S. Buist, P. M. Calverley, C. R. Jenkins, and S. S. Hurd, "Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease," *American journal of respiratory and critical care medicine*, 2012.

[7] M. Marschollek, "Decision support at home (ds@ home)–system architectures and requirements," *BMC medical informatics and decision making*, vol. 12, no. 1, p. 43, 2012.

[8] B. Arnrich, O. Mayora, J. Bardram, G. Tröster *et al.*, "Pervasive healthcare," *Methods of information in medicine*, vol. 49, no. 1, pp. 67–73, 2010.

[9] T. K. Liang, T. Tanaka, H. Nakamura, and A. Ishizaka, "A neural network based computer-aided diagnosis of emphysema using ct lung images," in *SICE, 2007 Annual Conference*. IEEE, 2007, pp. 703–709.

[10] V. Cheplygina, L. Sorensen, D. M. Tax, J. H. Pedersen, M. Loog, and M. de Bruijne, "Classification of copd with multiple instance learning," in *2014 22nd International Conference on Pattern Recognition (ICPR)*. IEEE, 2014, pp. 1508–1513.

[11] P. Arpaia, C. Manna, G. Montenero, and G. D'Addio, "In-time prognosis based on swarm intelligence for home-care monitoring: A case study on pulmonary disease," *Sensors Journal, IEEE*, vol. 12, no. 3, pp. 692–698, 2012.

[12] D. Newandee, S. Reisman, M. Bartels, and R. De Meersman, "Copd severity classification using principal component and cluster analysis on hrv parameters," in *Bioengineering Conference, 2003 IEEE 29th Annual, Proceedings of*. IEEE, 2003, pp. 134–135.

[13] C. C. Bellos, A. Papadopoulos, R. Rosso, and D. I. Fotiadis, "Identification of copd patientsâ€™ health status using an intelligent system in the chronious wearable platform," *Biomedical and Health Informatics, IEEE Journal of*, vol. 18, no. 3, pp. 731–738, 2014.

[14] G. Spina, P. Casale, P. S. Albert, J. Alison, J. Garcia-Aymerich, R. W. Costello, N. A. Hernandes, A. J. van Gestel, J. D. Leuppi, R. Mesquita *et al.*, "Identifying physical activity profiles in copd patients using topic models," *Biomedical and Health Informatics, IEEE Journal of*, vol. 19, no. 5, pp. 1567–1576, 2015.

[15] A. Agusti, J. Hedner, J. Marin, F. Barbé, M. Cazzola, and S. Rennard, "Night-time symptoms: a forgotten dimension of copd," *European Respiratory Review*, vol. 20, no. 121, pp. 183–194, 2011.

[16] R. A. Kinsman, R. Yaroush, E. Fernandez, J. F. Dirks, M. Schocket, and J. Fukuhara, "Symptoms and experiences in chronic bronchitis and emphysema." *CHEST Journal*, vol. 83, no. 5, pp. 755–761, 1983.

[17] T. Lee-Chiong, "Chronic obstructive pulmonary disease and sleep," *Current Respiratory Care Reports*, vol. 2, no. 2, pp. 123–129, 2013.

[18] S. Ancoli-Israel, R. Cole, C. Alessi, M. Chambers, W. Moorcroft, and C. Pollak, "The role of actigraphy in the study of sleep and circadian rhythms. american academy of sleep medicine review paper," *Sleep*, vol. 26, no. 3, pp. 342–392, 2003.

[19] D. M. Nunes, V. M. de Bruin, F. M. Louzada, C. A. Peixoto, A. G. Cavalcante, C. Castro-Silva, and P. F. de Bruin, "Actigraphic assessment of sleep in chronic obstructive pulmonary disease," *Sleep and Breathing*, vol. 17, no. 1, pp. 125–132, 2013.

[20] J. Hartman, J. Prinzen, R. van Lummel, and N. ten Hacken, "Frequent sputum production is associated with disturbed nightâ€™s rest and impaired sleep quality in patients with copd," *Sleep and Breathing*, vol. 19, no. 4, pp. 1125–1133, 2015.

[21] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *the Journal of machine Learning research*, vol. 3, pp. 993–1022, 2003.

[22] M. Jette, K. Sidney, and G. Blümchen, "Metabolic equivalents (mets) in exercise testing, exercise prescription, and evaluation of functional capacity," *Clinical cardiology*, vol. 13, no. 8, pp. 555–565, 1990.

[23] "BodyMedia SenseWear," http://sensewear.bodymedia.com/, 2016, [Online; accessed 16-January-2016].

[24] M. M. Sharif and A. S. BaHammam, "Sleep estimation using bodymedia's sensewearâ„¢ armband in patients with obstructive sleep apnea," *Annals of thoracic medicine*, vol. 8, no. 1, p. 53, 2013.

[25] H. Watz, B. Waschki, T. Meyer, and H. Magnussen, "Physical activity in patients with copd," *European Respiratory Journal*, vol. 33, no. 2, pp. 262–272, 2009.

[26] T. Troosters, F. Sciurba, S. Battaglia, D. Langer, S. R. Valluri, L. Martino, R. Benzo, D. Andre, I. Weisman, and M. Decramer, "Physical inactivity in patients with copd, a controlled multi-center pilot-study," *Respiratory medicine*, vol. 104, no. 7, pp. 1005–1011, 2010.

[27] H. Al Otair, M. Al-shamiri, M. Bahobail, M. M. Sharif, and A. S. BaHammam, "Assessment of sleep patterns, energy expenditure and circadian

rhythms of skin temperature in patients with acute coronary syndrome," *Medical science monitor: international medical journal of experimental and clinical research*, vol. 17, no. 7, p. CR397, 2011.

[28] J. Lin, E. Keogh, S. Lonardi, and B. Chiu, "A symbolic representation of time series, with implications for streaming algorithms," in *Proceedings of the 8th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery*. ACM, 2003, pp. 2–11.

[29] S. P. Lloyd, "Least squares quantization in pcm," *Information Theory, IEEE Transactions on*, vol. 28, no. 2, pp. 129–137, 1982.

[30] D. M. Blei and J. D. Lafferty, "Topic models," *Text mining: classification, clustering, and applications*, vol. 10, no. 71, p. 34, 2009.

[31] T. Huynh, M. Fritz, and B. Schiele, "Discovery of activity patterns using topic models," in *Proceedings of the 10th international conference on Ubiquitous computing*. ACM, 2008, pp. 10–19.

[32] Y. Rubner, C. Tomasi, and L. J. Guibas, "The earth mover's distance as a metric for image retrieval," *International journal of computer vision*, vol. 40, no. 2, pp. 99–121, 2000.

[33] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[34] J. H. Goodnight, "Tests of hypotheses in fixed effects linear models," *Communications in Statistics-Theory and Methods*, vol. 9, no. 2, pp. 167–180, 1980.

[35] D. Bates, M. Maechler, B. Bolker, and S. Walker, "lme4: Linear mixed-effects models using eigen and s4," *R package version*, vol. 1, no. 4, 2013.