RESEARCH ARTICLE

Open Access

Comparative analysis of predictive methods for early assessment of compliance with continuous positive airway pressure therapy

Xavier Rafael-Palou[1,2]* , Cecilia Turino[3,4], Alexander Steblin[1], Manuel Sánchez-de-la-Torre[3,4], Ferran Barbé[3,4] and Eloisa Vargiu[1]

Abstract

Background: Patients suffering obstructive sleep apnea are mainly treated with continuous positive airway pressure (CPAP). Although it is a highly effective treatment, compliance with this therapy is problematic to achieve with serious consequences for the patients' health. Unfortunately, there is a clear lack of clinical analytical tools to support the early prediction of compliant patients.

Methods: This work intends to take a further step in this direction by building compliance classifiers with CPAP therapy at three different moments of the patient follow-up, before the therapy starts (baseline) and at months 1 and 3 after the baseline.

Results: Results of the clinical trial shows that month 3 was the time-point with the most accurate classifier reaching an f1-score of 87% and 84% in cross-validation and test. At month 1, performances were almost as high as in month 3 with 82% and 84% of f1-score. At baseline, where no information of patients' CPAP use was given yet, the best classifier achieved 73% and 76% of f1-score in cross-validation and test set respectively. Subsequent analyzes carried out with the best classifiers of each time point revealed baseline factors (i.e. headaches, psychological symptoms, arterial hypertension and EuroQol visual analog scale) closely related to the prediction of compliance independently of the time-point. In addition, among the variables taken only during the follow-up of the patients, Epworth and the average nighttime hours were the most important to predict compliance with CPAP.

Conclusions: Best classifiers reported high performances after one month of treatment, being the third month when significant differences were achieved with respect to the baseline. Four baseline variables were reported relevant for the prediction of compliance with CPAP at each time-point. Two characteristics more were also highlighted for the prediction of compliance at months 1 and 3.

Trial registration: ClinicalTrials.gov Identifier, NCT03116958. Retrospectively registered on 17 April 2017.

Keywords: Obtrusive sleep apnea, Continuous positive airway pressure, Predictive methods, Machine learning

*Correspondence: xavier.rafael@eurecat.org
[1]Eurecat Centre Tecnòlogic de Catalunya, eHealt Unit, Carrrer Bilbao, 72, 08005 Barcelona, Spain
[2]BCN Medtech, Department of Information and Communication Technologies, Universitat Pompeu Fabra, Barcelona, Spain
Full list of author information is available at the end of the article

Background

Obstructive sleep apnea (OSA) [1] is defined as repeated episodes of shallow or paused breathing during sleep, despite the effort to breathe. This syndrome is caused by complete or partial obstructions of the upper airway leading to daytime sleepiness and impaired cardiopulmonary function. Gold standard treatment of OSA involves the use of a device that administers continuous positive pressure (CPAP) in the respiratory tract of patients [2]. Patients should wear a mask, usually nasal [3, 4], to receive the air from the CPAP device. Doctors are responsible for adjusting the air pressure provided by the device, preventing the throat from collapsing or damaging its tissue [5].

Several studies aimed at objectively assess the effects of using CPAP show that this kind of intervention is highly effective in improving symptoms, such as daytime sleepiness, morbidity, and mortality rates related to cardiovascular diseases [6, 7]. Although it is highly effective in minimizing OSA symptoms, up to 36% of patients do not use or even discontinue CPAP [8, 9].

Different factors have been found that influence the adherence to CPAP treatment, although they have not been reported consistently. In [10] patient demographic features, such as age, sex, and marital status were found to be weak predictors of CPAP adherence, as well as apnea-hypopnea index (AHI), oxygen desaturation index (ODI), or daytime sleepiness in [11–13]. In contrast, an improvement of ODI and deep sleep in patients during the CPAP titration increased the chances of complying with CPAP in [14]. Also, the adherence measured few days after CPAP initiation was shown to be a good predictor of long-term compliance in [15].

Despite these efforts, the reasons that lead to compliance with therapy remain an open field of research. In fact, knowing in more detail the initial factors that determine compliance with CPAP could help reduce dropout rates and the application of personalized treatments to improve patient compliance. Unfortunately, there is a clear lack of clinical analytical tools to support the early prediction of compliant patients.

When statistically analyzing the possible factors that might determine CPAP compliance, we found common complexities that compromise the predictive capacity and robustness of the findings. The limited number of participants, the large number of clinical variables and the quality of collected data are just to name a few. To overcome these limitations the use of machine learning (ML) might be an alternative solution to the aforementioned problems. Despite the numerous ML applications in the medical domain, (e.g. disease diagnosis [16, 17]), compliance with therapy is usually constrained to the medication adherence problem [18, 19]. Recent ML algorithms (e.g. support vector machines [20] and artificial neural networks [21]) often provide highly accurate predictive

models. However, such models lack transparency and therefore their interpretation is difficult [22]. As a consequence, other classification algorithms in the medical field are still preferred (e.g. logistic regressions [23] or decision trees [24]).

The goal of the study presented in this paper is twofold. On the one hand, to provide a comparative analysis of predictive methods of CPAP compliance built using machine learning techniques in different stages of treatment. On the other hand, to define the most important factors associated with CPAP compliance identified by the best predictive methods obtained in the different stages of therapy up to month 6.

Methods

Participants

Fifty-one adult patients (> 18 years), diagnosed with OSA (15 or more apneas/hypopneas per hour in an overnight sleep study) and requiring CPAP treatment were recruited at Hospital Arnau de Vilanova (Lleida, Spain). Patients with impaired lung function (overlap syndrome, obesity hypoventilation, and restrictive disorders), severe heart failure, psychiatric disorders, periodic leg movements, pregnancy, other dyssomnias or parasomnias, and/or a history of previous CPAP treatment were excluded. The study was approved by the hospital ethics committee (Approval number: CEIC-1283). All recruited patients signed an informed consent form.

Of the 51 patients originally included in the study, 3 were excluded due to malfunction of the CPAP machine, 5 did not attend the last visit at the sleep unit and 1 patient died during the study.

The final sample consisted of 42 patients (29 males and 13 females) with a mean age of 56.93+/-12.58 yrs. Their BMI was 33.83+/-6.46 and their number of apnea or hypopneas per hour of sleep (apnea/hypopnea index or AHI) was 53.13+/-20.72 events/h. In our sample, 60% (25) of all patients were active workers and 33% (14) were retired. The sample also had 62% of nonsmokers (26) and 57% of non-alcoholic consumers (24). In terms of CPAP device use, the patients scored an average of nightly hours of use of 5.44+/-1.74 at month-1, 5.33+/-1.90 at month-3, 5.07+/-2.10 at month-6.

Datasets

The study variables from the 42 patients were manually collected by lung specialists along four visits at month-0 (baseline or T0), at month-1 (T1), at month-3 (T3), and at month-6 (T6). During the first visit (T0) clinicians gathered 77 features organized in five categories: clinical history (e.g. depression, anxiety, arterial hypertension (HTA), cardiopathy, neurological disease, respiratory disease), symptoms (e.g. irritability, apathy, depression, insomnia), co-morbidities (e.g. diabetes, obesity, dyslipidemia),

Rafael-Palou *et al. BMC Medical Informatics and Decision Making*   (2018) 18:81

Page 3 of 14

therapies (e.g. beta blockers, diuretics), sleeping test (e.g. sleeping time, AHI, percentage of the night spent with oxygen saturation < 90% or CT 90) and basal information (e.g. size, weight, BMI, tas, tad, oxygen saturation). In the second visit (T1), after the patients had the CPAP machine at home during one consecutive month, 16 new features related to monitoring were collected (e.g. nightly average use, abandon or adverse effects of the treatment, such as dry mouth, allergies, and cutaneous irritations). At the third month (T3), the same number of features as in T1 were gathered together with 5 more: size, weight, BMI, removed drugs, and added drugs). At month-6 (T6), although some other variables were collected, for the purpose of this study only the average use of nightly hours was considered. Eventually, three datasets (D0, D1, and D3) with an incremental number of features (i.e. D1 features = D0 features + features collected at T1) were created with 77, 93, and 114 features, respectively. The full list of variables is described in Additional file 1: Table S1 of the supplementary material.

In this study, we addressed CPAP compliant users as those who had more than 4 h per night on average during the first 6 months of treatment. Therefore, all samples from each dataset (ds) were labeled using the collected information about nightly hours/use on average of the CPAP device at the end of the month-6. In so doing, 24 (57%) patients were labeled as "compliant", class "1", as they correctly followed the CPAP therapy prescription (more than 4h nightly on average). On the contrary, 18 (43%) patients did not achieve the prescribed treatment (minus or equal than 4 h nightly on average) and they were labeled as "non-compliant", class "0".

Data description

Datasets D0, D1, and D3 collected at time points T0, T1, and T3, respectively, were statistically analyzed for a better understanding of the sample. In this task, the Mann-Whitney U test was used to evaluate the statistical significance of quantitative variables with CPAP compliance and Chi-square tests for qualitative characteristics. Previously, the categorical features were converted into numeric to achieve a homogeneous data type sample. Variables with only two categories were directly mapped into binary values. Variables with more than two categories, given their underlying incremental meaning, were mapped into unsigned ordinal values.

Data cleansing

We carried out a set of empirical tasks, supervised by the lung specialists, to reduce possible noise and redundancy in the datasets. First, given the existence of null values in the datasets, an imputation process was carried out consisting of computing the mode for the categorical characteristics and the mean for the numerical characteristics.

Subsequently, the distributions of the categorical characteristics were analyzed, which revealed variables with few individuals by category. The features with a number of individuals less than a given threshold were removed from the study to avoid the noise they might introduce when building the predictive models. To catch up possible information redundancy in the datasets, we computed the mutual information (MI) [25] score for the categorical variables in a pair-wise manner. From each pair, we kept the variable more statistically significant with the dependent variable using Chi-square test. Among the numerical features, we applied a correlation analysis to detect highly redundant features. Given the existence of non-normally distributed numerical features, we used the Spearman correlation method on all numerical variables in a pair-wise manner. Empirically we set-up a threshold for the correlation scores above of which one feature of the pair was removed (i.e. the feature with the highest *P*-value). *P*-values lower than 0.05 were considered statistically significant.

Classification framework

All preprocessed datasets presented common particularities, such as a small number of samples, the presence of missing values, class unbalance and high multidimensionality feature space. To cope with these complexities we designed a classification framework flexible enough to enable the execution of heterogeneous pipelines or sequence of configurable machine learning steps. In particular, the pipelines were composed of three mandatory steps (i.e. imputation, variance filtering and data standardization), two optional steps (i.e. feature selection and feature sampling), and two more final steps (i.e. classifier training and evaluation). In total 80 pipelines were configured from 4 feature selection methods, 5 classifier algorithms, 2 sampling strategies, and 2 evaluation metrics.

The result of running (i.e. training or building) a pipeline (Pipe$_i$) on a dataset (D_j) with parameters (*params$_i$*) is a predictive model or classifier ($M_{i,j}$) with its associated predictive performance (Perf$_{i,j}$). Figure 1 sketches the scheme with the input, output, and the different steps that configure the pipelines for compliance with CPAP therapy.

The first step of a pipeline is the imputation of null values. To do this, given the small proportion of null values in the datasets, a simple strategy was proposed to replace the null values with their most frequent value (for categorical characteristics) and with the mean value (for numerical characteristics).

The second step consists of a simple filter method to eliminate features with zero variance, that is, to eliminate these characteristics that have the same value in all the samples and that do not provide any additional information to the dataset.
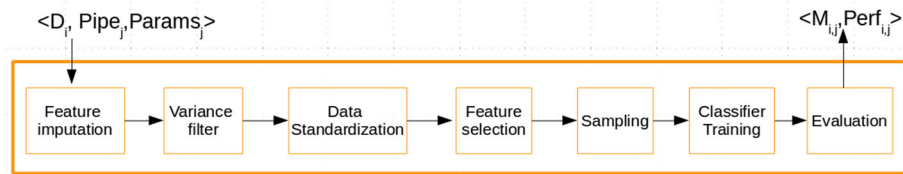
Rafael-Palou *et al. BMC Medical Informatics and Decision Making* (2018) 18:81

Page 4 of 14

Fig. 1 Pipeline steps designed for building classifiers for compliance with the CPAP therapy

Since the data come from different sources, the next step is to standardize them. This step consisted of homogenizing all features to zero mean and variance one. This transformation step is crucial for the construction of many classification algorithms since it allows them to compare features without harming their performance or execution time [26].

Feature selection (fs) was introduced in the pipelines given a large number of features compared with the number of samples for each dataset ($p > n$). This type of methods aims to reduce over-fitting by improving model performance and generalization, to provide faster and more cost-effective models and simplify models making them easier to interpret [27]. Feature selection methods are usually divided into three categories: filter, wrapper, and embedded [28]. Filter methods, in general, examine features individually with respect to the class, wrapper methods use a predictive model to generate subsets of features evaluated according to their predictive power, and embedded methods search for an optimal subset of features during the training of the prediction model. In this study, we used one method for each of the different feature selection strategies. For the filter-based strategy, we defined a simple method (Combine_fs) that makes a ranking of the features by their statistical significance with the class (i.e. applying ANOVA or chi-squared tests according to the data type of the characteristics). Then, this method returns the subset of features through a configurable threshold. For the wrapper strategy, we proposed the recursive feature elimination (RFE_RF_fs) method [29] configured with a random forest to provide the importance of the features. The embedded strategy was in charge of the application of the least absolute shrinkage and selection operator (Lasso_fs) [30]. To this purpose, we used a logistic regression model configured with the L1 norm. It was also considered the possibility of not using any method of feature selection.

The next step is data sampling (sm). In particular, we proposed the use of the synthetic minority over-sampling technique (Smote) [31]. This technique consists of creating synthetic samples (i.e. detecting similar instances and performing small perturbations in their values) of the under-represented class samples instead of creating copies, as the over-sampling method would do. In particular, we used this technique to balance the minority class with the same number of instances of the majority class. The main idea behind this method is to avoid the bias produced by many standard classifier learning algorithms towards the class with a larger number of instances. As in the feature selection pipeline step, we also considered the option of not using data sampling in the experiments.

Regarding the training and evaluation stage, we selected several classification algorithms (cls) to deal with various classification strategies (i.e. linear, non-linear, distance-based, and tree-based). In fact, the provision of different classification strategies is especially appropriate in complex datasets when the distribution of data is not clear. As already mentioned, the interpretability of the resulting predictive models is also a desired condition. Therefore, we opted for logistic regression (LR) [23], k-nearest neighbor (k-NN) [32], and random forest (RF) [33] for the subset of interpretable classification algorithms (referred as "descriptive", hereinafter). In contrast, we chose support vector machines (SVM) [20] and artificial neural networks (NN) [21] for the subset of algorithms with less interpretative capacity but with a potential greater discriminatory capacity (referred as "non-descriptive", hereinafter).

Evaluation setup

In order to ensure adequate performance evaluation, the available data were shuffled, stratified, and randomly divided into train (29 rows, 12 non-compliant, and 17 compliant) and test (13 rows, 6 non-compliant, and 7 compliant) sets with a ratio of 70/30. Therefore, the training set partitions of the three datasets (D0, D1, and D3) contained the same individuals. The same rule applies for the test set.

Test sets remained untouched until the end of the process. Training sets were used for 10-fold cross-validation to enable proper model tuning and evaluation. This technique is particularly suitable when the sample size is small. Indeed, as suggested in [34], the entire sequence of processes that composed each pipeline was wrapped-up within the cross-validation technique in order to reduce the possibility of biased results. Thus, training data were shuffled and randomly split into stratified train-validation

Rafael-Palou *et al. BMC Medical Informatics and Decision Making* (2018) 18:81

Page 5 of 14

sets (20 rows, 8 non-compliant, and 12 compliant) and stratified test-validation sets (9 rows, 4 non-compliant, and 5 compliant) following a ratio of 70/30. Then, for each of the configured pipelines (i.e. 80 pipelines), we created as many experiments as combinations of values for the different hyper-parameters defined for each method of the pipeline (Table 1).

We performed 10-fold cross-validation for all experiments in each pipeline. This process was repeated for each of the proposed learning metrics (i.e. f1-weighted and precision-weighted). The learning metric, f1-weighted

Table 1 Pipeline parameters tested using grid-search and 10-fold CV

Pipeline step	Parameter options
Combine_fs	percentile = [5, 10, 20, 30, 40, 50]
Lasso_fs	estimator = Logistic Regression
	penalty = "l1"
	$C =$ [5, 10, 20, 30, 40, 50]
RFE_RF_fs	class_weight = 'balanced'
	n_estimators = 100
	step = [0,1]
	n_features_to_select = [0.4,0.6,0.8]
Smote_fs	n_neighbors = [3,4,5]
	ratio='auto'
	kind='regular'
k-NN	n_neighbors = [1,3,5,7,9,11]
	weights = ['uniform', 'distance']
LR	$C =$ [0.00001, 0.0001, 0.0005, 0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 1, 5, 10, 15, 30]
	class_weight = [None, 'balanced']
	penalty = ['l1', 'l2']
RF	n_estimators = [100,150,200,250,500]
	criterion = ['entropy','gini']
	max_depth = ['None',4,6]
	class_weight = [None, 'balanced']
SVM	$C =$ [0.01,0.1,0.5,1,5,10,15,30,50]
	gamma = [0.0001,0.001,0.01, 0.1,1,5]
	kernel = 'radial'
	class_weight = [None, 'balanced']
NN	alpha = [1e−5, 0.00001, 0.0001, 0.001, 0.01,0.1,1,3,5,10]
	hidden_layer_sizes = [(30,), (50,), (70,), (100,), (150,),
	(30,30),(50,50),(70,70),(100,100),
	(30,30,30),(50,50,50),(70,70,70)]

($f1$), was selected since it is a suitable measure for unbalanced datasets. This metric combines the precision and recall metrics weighted by the number of samples per each class. The other selected metric, precision-weighted (*prec*), tends to prefer classifiers with less incorrect compliant predicted patients, which indeed they are the most harmful cases to avoid. This metric computes the ratio of correctly classified cases (i.e. compliant patients) among all positive classified cases weighted by the number of samples per each class.

As a result of this 10-fold cross-validation process, we reported for each experiment the average and standard deviation performance of the learning metric which the pipeline was configured with. A greedy-search strategy was applied to select the best experiment, i.e. best pipeline parameterization. Additionally, with the intention to avoid the possible bias introduced in this process to the validation dataset [35] and to reduce the chances of having obtained similar folds with repeated individuals, we evaluated the best parameterization of each pipeline (i.e. pipelines with the appropriated values for their hyper-parameters) using a final outer stratified 10-fold cross-validation on the training data (i.e. with learning metric=f1-weighted, ratio=70/30 for cv-train and cv-test). As a result of this process we reported the final cross-validation performance (f1-score) of the pipelines. This score was provided by the f1-weighted metric since although having a high precision is desirable, a high recall (*rec*) is also needed especially for health institutions since it reduces false negatives and thus non-necessary clinical interventions and additional costs.

This whole process was repeated 80 times for all pipelines of each dataset. The best pipelines of each dataset were identified by ranking the cross-validation performances (i.e. f1-score) reported by each pipeline. The best pipelines of each dataset were compared together in order to find statistically significant differences. We did the same among the best descriptive and non-descriptive pipelines. To do this, we used a 10-fold cross-validated paired t-test.

To complete the reporting of this analysis, we computed on the test set and for each of the best pipelines of each dataset a comprehensive set of scores (i.e. f1, precision, recall, AUC, and confusion matrix) to enable a better understanding of the results. Also, we reported their ROC and learning curves.

Feature importances
We reported the most important features from the best descriptive pipelines. Let us note that, in this study, we only used those configured with random forest and logistic regression. for each dataset. To do this, we performed a ranking of features using a stability score [36]. This score measures how "stable" the features of a predictive model

Rafael-Palou *et al. BMC Medical Informatics and Decision Making* (2018) 18:81

Page 6 of 14

are. For that reason, we build a pipeline n times using n random subsets of fixed size s.

To compute this score, we created $n = 100$ randomly stratified partitions from the $s = 70\%$ of the entire dataset. For each data partition, we trained the best pipelines and keep a record of the selected features of the classifier and their weights (or feature importances for RF classifier). With this information, we computed the number of times any feature was selected (i.e. stability score) and its normalized absolute average weight (or importance). Since one classifier might report all features as relevant (i.e. non-zero), a threshold ($t > 0.4$) in the weights was empirically defined to make usable the stability score.

Results

Data cleansing

A descriptive analysis of the initial datasets was carried out and summarized in Tables S2, S3 and S4 of the Additional file 1. Subsequently, the data cleansing process was conducted under the supervision of the lung specialists. From this process, we found out that (11/27/42) features had null values with ratios between 2.3% and 12% from the total number of rows. After a null imputation, we found (14/7/10) variables with underrepresented categories ($<= 10\%$ of rows per category). We also detected 4 pairs of categorical features (i.e. no active, anti-depressives, ADO, and memory disorders) in the D0 with MI scores above 50%. From the correlation analysis applied on the numerical variables, we found 4 highly ($> 80\%$) redundant features in D0 (i.e. abdomen and hip circumference, weight and CT90%) and 4 features in D3 (i.e. size_3, weight_3, bmi_3 and total_use_hours_3). After the validation by the specialists of setting aside the aforementioned features, we obtained the final datasets for the classification analysis. These were composed of (54/63/70) variables, respectively.

Classification analysis

We evaluated 76 out of 80 initially configured pipelines for each dataset. In particular, we rule out pipelines which had same classification algorithm (i.e. random forest) for

feature selection and classification given their initial poor contribution to the experimental results and the long runtime required to complete their evaluation.
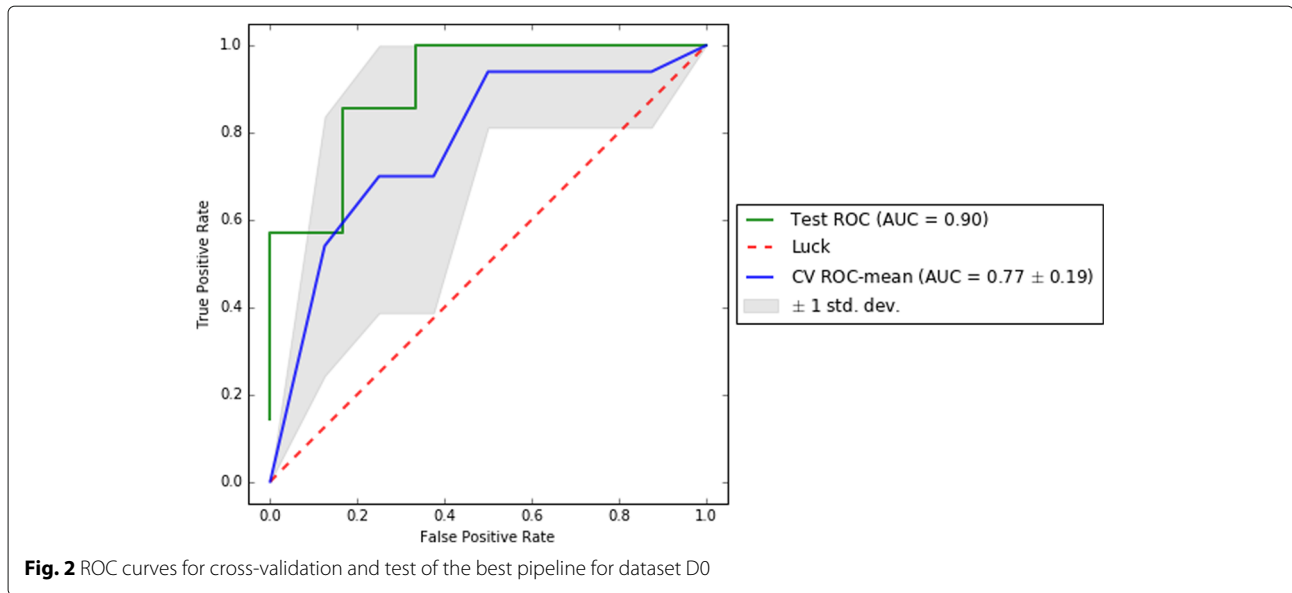
Best pipelines (p0, p1, p3) for D0, D1, and D3, respectively, achieved 0.73+/-0.18, 0.82+/-0.06, 0.87+/- 0.15 of f1-score in cross-validation and 0.76, 0.84, 0.84 in test set (Table 2). These pipelines were configured with precision-weighted metric and SVM algorithm for the D0 dataset; with Smote sampling, f1-weighted metric, and an SVM for the D1 dataset, and with Lasso feature selection, Smote sampling, precision-weighted metric and an RF for the D3 dataset.

To visually support these values, Figs. 2, 3, and 4 show the area under the receiver operating characteristic (ROC) curves and Figs. 5, 6, and 7 show the learning curves with the effects of increasing the size of the training set in their performances.

We also analyzed which was the contribution in classification performance of the different techniques configured in the pipelines (i.e. sampling strategy, feature selection, learning metric, and classification algorithm). In particular, the average in the performance (i.e. f1-score) of all pipelines using the sampling strategy was (0.59+/-0.07,0.61+/-0.09,0.75+/-0.09) for each dataset. Let us recall that Smote [31] was the selected sampling technique to balance the training data from each fold of cross-validation by oversampling the minority class to the number of samples of the majority class (i.e. from 8 non-compliant/12 compliant to 12/12). On the contrary, not using any sampling strategy was (0.58+/-0.07,0.62+/-0.08,0.75+/-0.08). Concerning the average performance reached among the pipelines configured with the best feature selection methods they scored (0.59+/-0.03,0.63+/-0.07,0.77+/-0.08). On the other side, the average in performance reached by the pipelines without using any feature selection gave as result (0.66+/-0.07,0.68+/-0.11,0.77+/-0.09), respectively. Focusing on the evaluation metric with which the pipelines were configured, the pipelines with f1-weighted metric achieved an average in the performance of (0.58+/-0.07,0.62+/-0.09,0.76+/-0.09) while the pipelines configured with precision-weighted obtained an average performance of ($0.59 + / - 0.07, 0.61 + / -$

Table 2 Performances of the best pipelines in each dataset

id	ds	sm	fs	metric	cls	params
p0	D0	none	none	precision_weighted	SVM	[0.001, balanced, 30]
p1	D1	Smote	none	f1_weighted	SVM	[0.001, None, 4, 15]
p3	D3	Smote	Lasso_fs	precision_weighted	RF	[1, 250, gini, 4, None, None]

id	cv_prec	cv_rec	cv_f1	test_prec	test_rec	test_f1
p0	0.78+/-0.2	0.74+/-0.17	0.73+/-0.18	0.77	0.85	0.76
p1	0.84+/-0.06	0.82+/-0.05	0.82+/-0.06	0.85	0.88	0.84
p3	0.89+/-0.14	0.88+/-0.14	0.87+/-0.15	0.85	0.88	0.84

Rafael-Palou *et al. BMC Medical Informatics and Decision Making* (2018) 18:81

Page 7 of 14

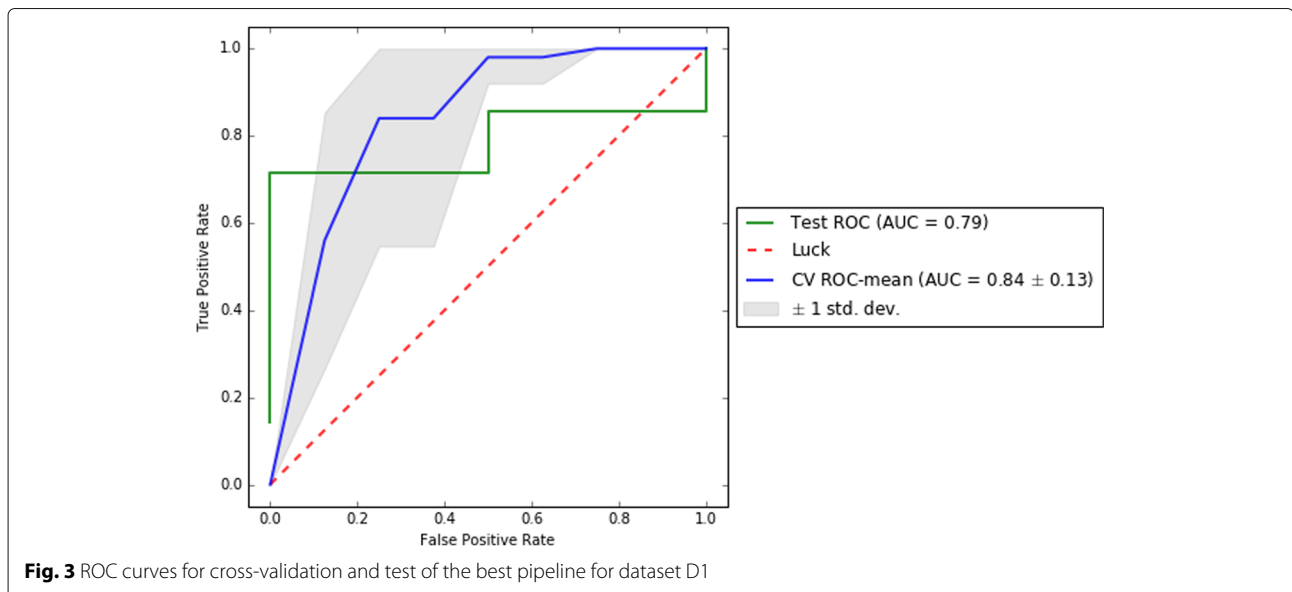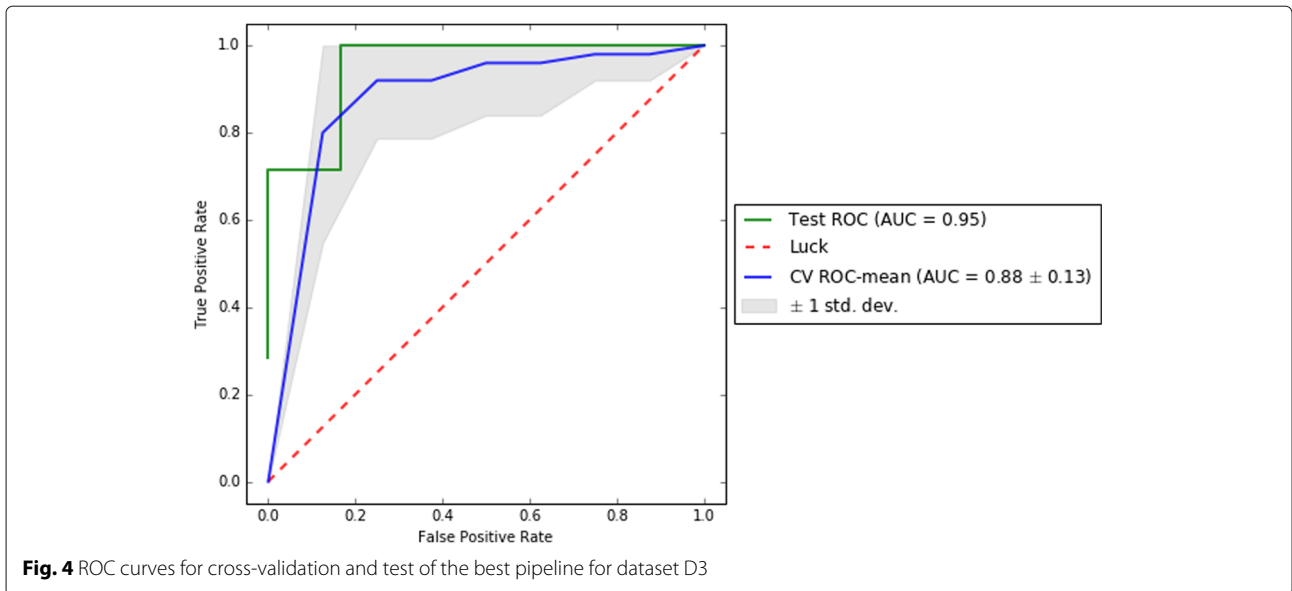**Fig. 2** ROC curves for cross-validation and test of the best pipeline for dataset D0

$0.09, 0.75 +/-0.09$). Regarding the type of classification algorithm (Fig. 8), the best pipelines reported performances in cross-validation between 0.59+/-0.21 (using k-NN) and 0.73+/-0.18 (using SVM) of f1-score in D0. In D1 the best pipelines reported performances between 0.61+/-0.12 (using k-NN) and 0.82+/-0.06 (with SVM). In D3 the best pipelines reported performances between 0.75+/-0.07 (using k-NN) and 0.87+/-0.15 (with RF). Table 3 summarizes these differences of performance among the configured pipelines.

Regarding the comparison of performance between descriptive and non-descriptive pipelines, the best descriptive pipelines obtained f1-scores of 0.69 +/-0.15, 0.75 +/- 0.15 and 0.87 +/- 0.15 in cross-validation and

0.76, 0.84, 0.84 in test set, while the best non-descriptive pipelines obtained scores of 0.73 +/-0.18, 0.82 +/- 0.06 and 0.84 +/- 0.08 in cross-validation and 0.76, 0.84, 0.84 in test set. Further details about these pipelines can be found in Table 4.

To complete the analysis, we extracted the most important features used by the best descriptive pipelines. In total (25/28/20) features were reported for each dataset after setting up a minimum threshold of 0.4 in the feature weights. Top-10 features of D0 and D1 reported stability scores above 0.6. In D3 only 2 features were above 0.4 of stability score. Figures 9, 10, and 11 provides a visual ranking of the features of each of the dataset ordered by the stability score.

Fig. 3 ROC curves for cross-validation and test of the best pipeline for dataset D1

Rafael-Palou *et al. BMC Medical Informatics and Decision Making*   (2018) 18:81

Page 8 of 14

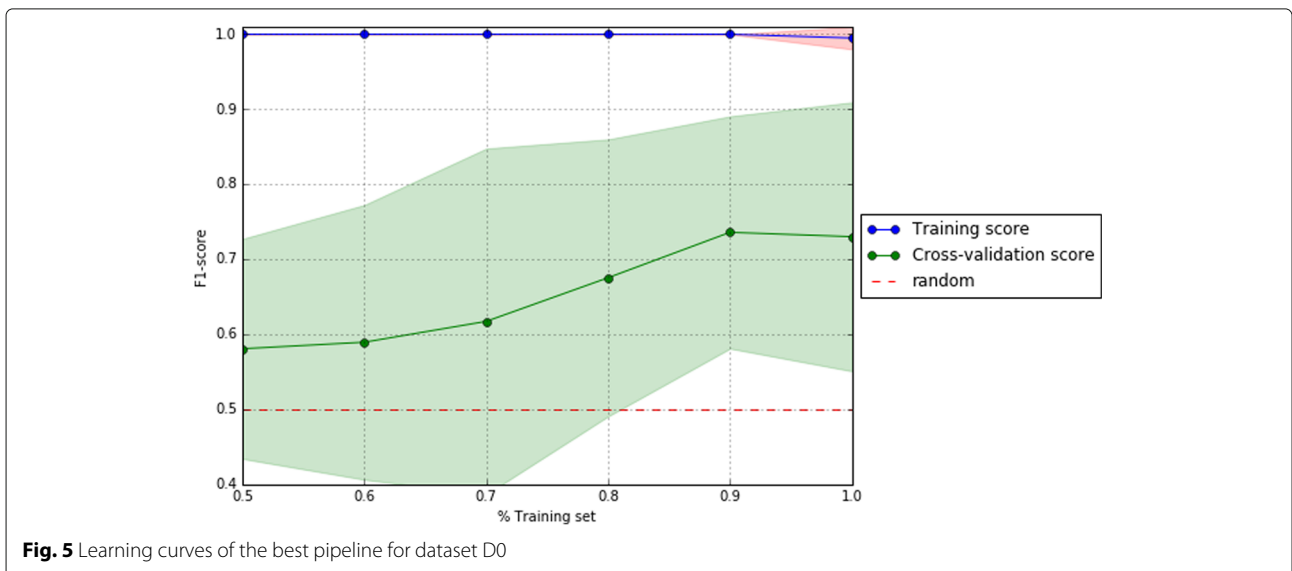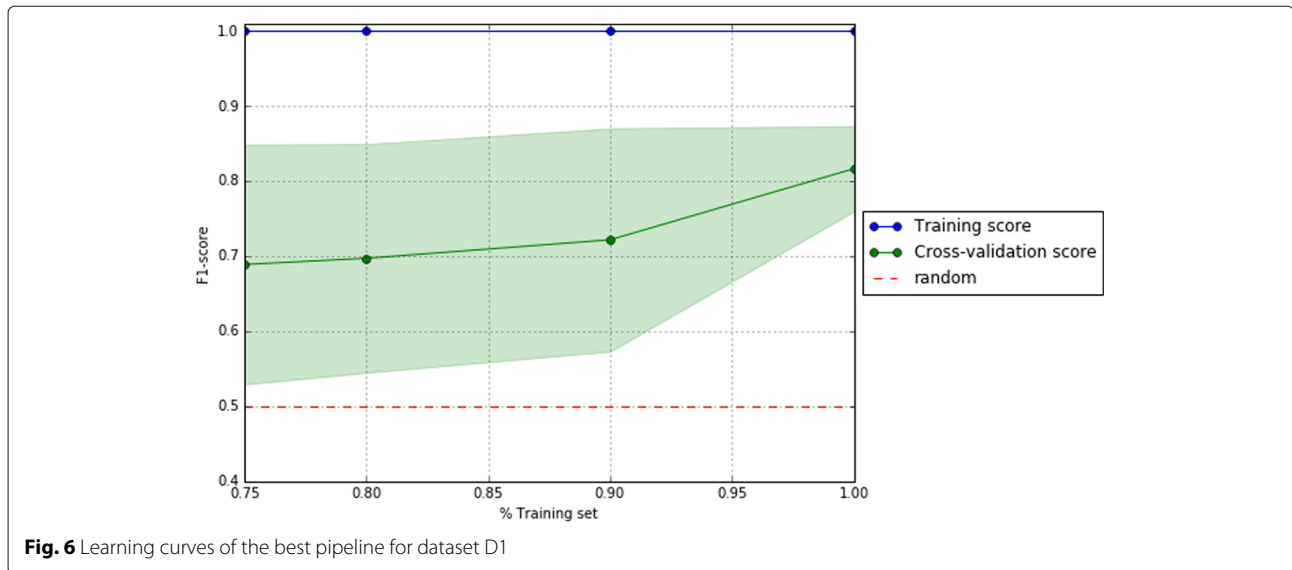**Fig. 4** ROC curves for cross-validation and test of the best pipeline for dataset D3

Discussion

CPAP compliance has been demonstrated to reduce cardiovascular risk and symptoms in patients with sleep apnea. Therefore, an automatic support to the early identification of patients with poor compliance with CPAP therapy would allow personalized treatments and better management of hospital resources. Consequently, in this article we proposed the creation of classifiers that predict compliance with CPAP therapy in patients with OSA in the early stages of treatment.

One of the main challenges of this study has been to build predictive models from small data [37]. In this scenario, models tend to over-fit training and even validation data, which means they are not able to generalize well for previously unseen data. To overcome this problem, first, under the supervision of pulmonary specialists, we performed an initial process of data cleansing to eliminate the non-relevant and noisy characteristics of the dataset. Secondly, we set up several machine learning experiments (e.g. using data sampling, feature selection or one of the different proposed machine learning algorithms) to obtain those with the best generalization performance for each dataset. To do this, we proposed the use of machine learning pipelines to carry out all the steps of the learning experiment together for each of the folds of the cross-validation. In this way, we avoided the knowledge leakage

Fig. 5 Learning curves of the best pipeline for dataset D0

Rafael-Palou *et al. BMC Medical Informatics and Decision Making* (2018) 18:81

Page 9 of 14

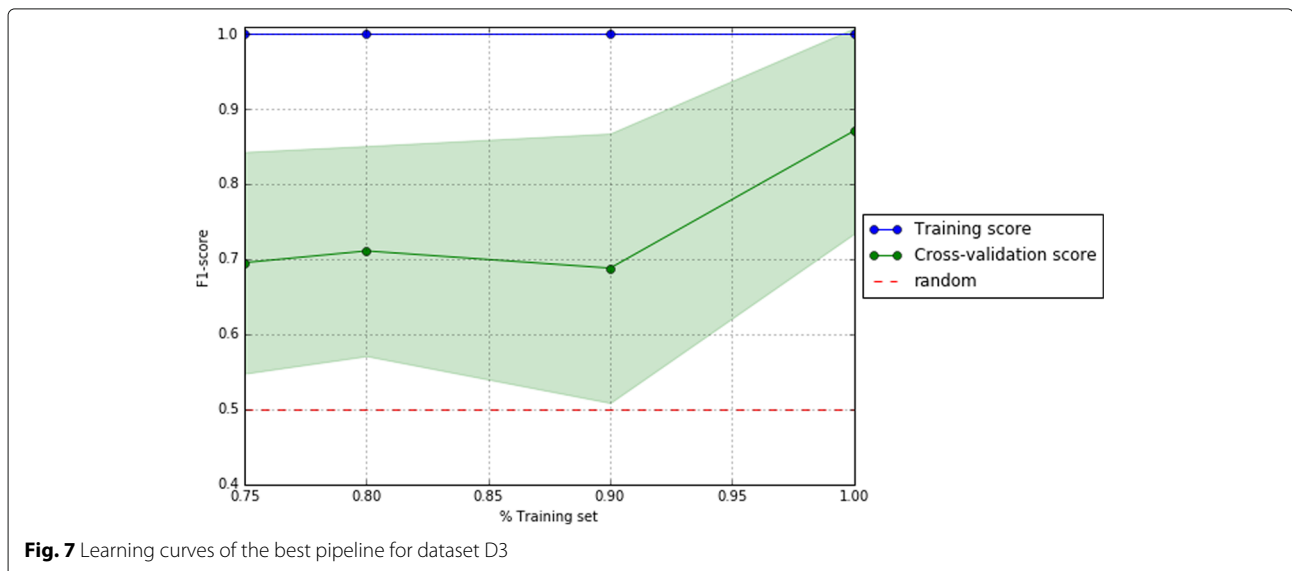**Fig. 6** Learning curves of the best pipeline for dataset D1

from the training to the evaluation stage. Thirdly, we used a nested cross-validation [38] to train, adjust, and evaluate the models. This mechanism reduces the chances of reporting over-fitted models since the validation scores are reported through a second level of cross-validation, once the models are tuned on the first level of cross-validation. Lastly, we provided a thorough evaluation of the best models to understand their performance. The results of this evaluation are presented in Table 2 and show that the performance of the best predictive models, both for validation and test datasets, are high (i.e. greater than 0.7 for D0 and higher than 0.8 for D1 and D3 of precision, recall, and f1 scores) and close to each other (i.e. with only 0.02 and 0.03 difference in f1) for each of the analyzed

datasets. These values reflect that current models are reliable and generalize well in data partitions that were not used for training. In addition, as can be seen in the Figs. 5, 6, and 7, validation curves reach the highest values when all training data are available to build the classifiers. However, as these are still a bit far from the training curves but the tendency is to get together, getting more data for training could be beneficial to obtain more accurate models (especially for datasets D1 and D3).

Going deep into the classification results obtained at three different moments of CPAP treatment, they show that dataset D0, collected before the start of treatment, is the most complex to learn compared to D1 and D3, collected in months 1 and 3 of the patient's therapy,
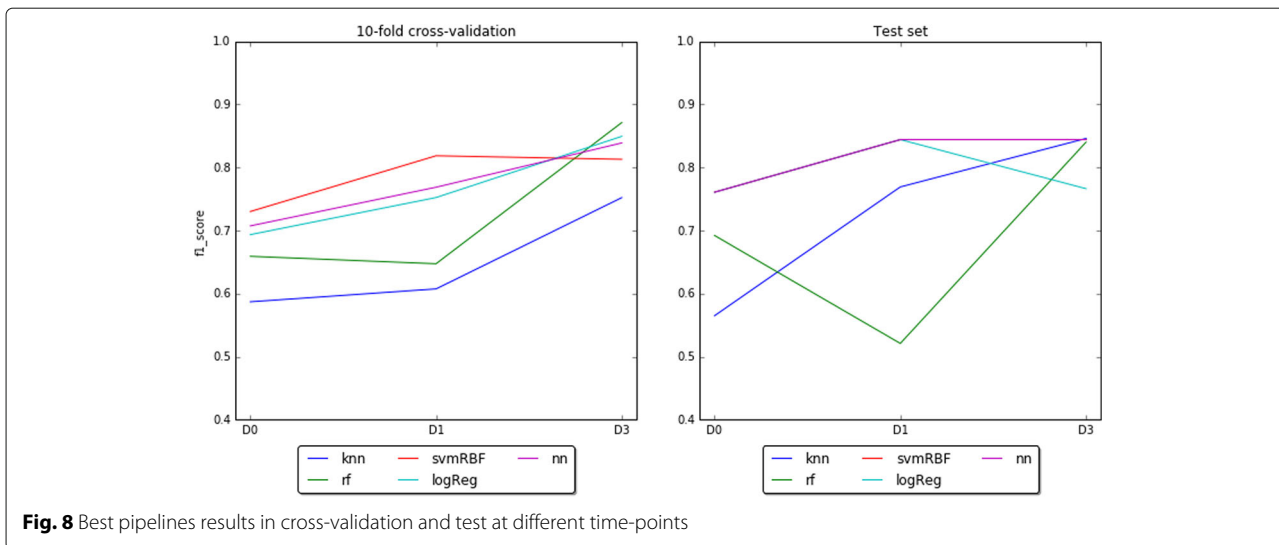

Fig. 7 Learning curves of the best pipeline for dataset D3

Rafael-Palou *et al. BMC Medical Informatics and Decision Making*   (2018) 18:81

Page 10 of 14

Fig. 8 Best pipelines results in cross-validation and test at different time-points

respectively. Nevertheless, a considerable average f1-score of 0.73 +/- 0.18 was achieved in cross-validation and 0.76 in the test set. As shown in Table 5, an important performance increase of 0.09 ($p = 0.12$) was reached in D1 with respect to D0. In D3, we get the best classification performance with a significant increase of 0.14 in f1-score ($p = 0.024$) with respect to D0. This same trend occurred in the test set where the best pipelines of D1 and D3 reported performances of 0.18 above the achieved in D0. The difference in f1-score between D3 and D1 did not prove to be significant ($p = 0.30$). These results seem to confirm that follow-up measurements help to increase baseline prediction performance. Indeed, the closer to the CPAP compliance cut-off we are the more confident the classifier is (i.e. performance in D0 is lower than performance in D1 and lower than D3). In addition, patterns of CPAP adherence appear early, in our case at month 1, since it is when the greatest increase in performance is achieved. This same finding was confirmed by other studies [15, 39].

During the evaluation step, we noticed that the use of sampling, feature selection or a particular learning metric
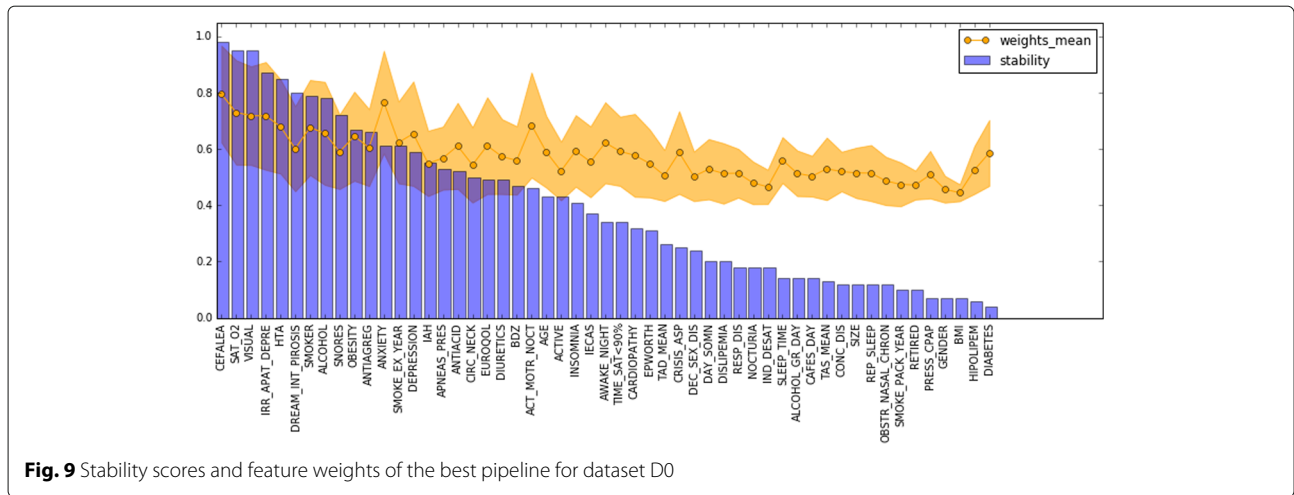
were not as substantial as expected in any of the datasets (Table 3). To be more specific, the maximum increase in performance, regardless of the method used, was between 0.02 and 0.04 of f1-score in cross-validation. Probably this confined contribution was due to the initial preprocessing and the fact that the data were not severely unbalanced. Indeed, in D0 and D1 the use of feature selection compromised the performances with a maximum decrease of 0.13. In contrast, important increments of performance in all datasets were produced depending on the classification algorithm used (i.e. 0.14, 0.21, and 0.20). In D0 and D1 best pipelines were using an SVM. This result was not surprising because this algorithm has been already found suitable for problems with few samples and with a high number of features [40], being able to build complex non-linear decision boundaries. In D3, the best pipeline was configured with an RF, although the one with SVM also provided a high score. RF algorithm is also suitable for difficult problems and especially indicated for handling categorical features [33]. The other non-descriptive classifier (i.e. NN) reported competent performances in all three datasets, especially in D3, where it exceeded the results reported

Table 3 Performance difference of f1 cross-validation along the different datasets achieved among pipelines configured with different techniques

Methods	Comparisons	D0		D1		D3	
		Avg	Max	Avg	Max	Avg	Max
Sampling	Smote vs None	0.01	-0.02	-0.01	0.04	0.0	0.02
Feature selection	Best vs None	-0.07	-0.07	-0.05	-0.13	0.0	0.02
Metrics	f1 vs prec	0.01	0.02	0.01	0.04	0.01	-0.01
Classifier algorithm	Best vs Worst	0.09	0.14	0.14	0.21	0.19	0.20

Table 4 Best descriptive and non-descriptive pipelines by dataset

ds	sm	fs	metric	cls	params
D0	none	none	prec	SVM	[0.001, balanced, 30]
D0	Smote	none	prec	LR	[None, 15, 4, l2]
D1	Smote	none	f1	SVM	[0.001, None, 4, 15]
D1	none	none	f1	LR	[None, 5, l2]
D3	Smote	Lasso_fs	prec	RF	[1, 250, gini, 4, None, None]
D3	none	none	f1	LR	[None, 0.5, l1]

Rafael-Palou *et al. BMC Medical Informatics and Decision Making* (2018) 18:81

Page 11 of 14

Fig. 9 Stability scores and feature weights of the best pipeline for dataset D0
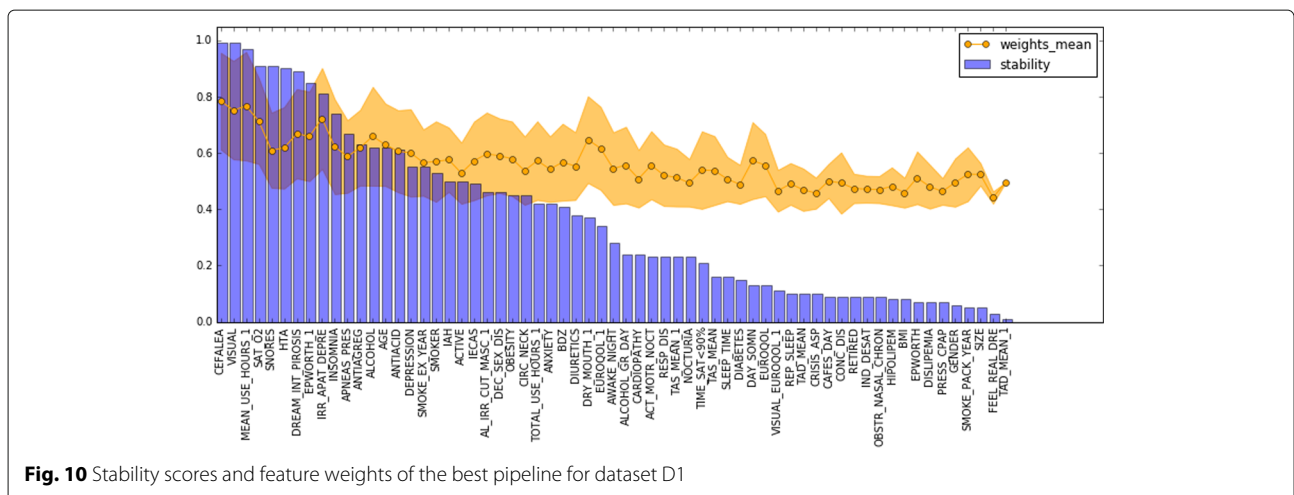
by the best pipeline configured with an SVM. However, the k-NN algorithm reported the worst scores of the three datasets. This is partly because it does not usually work well with a large number of features [41].

Focusing on the differences of performance between the best descriptive and non-descriptive pipelines, those were always lower than 0.1 and not significant ($p = 0.14$) for D0 but significant in D1 ($p = 0.02$). In contrast, the best performance in D3 was achieved through a descriptive classifier although in cross-validation the difference in performance with the best non-descriptive pipeline ($0.02+/-0.16$ of f1-score) proved to be non-significant ($p = 0.69$).

Regarding the relevant factors related to the CPAP compliance prediction, four baseline features were found common in each of the best pipelines for the different datasets collected at time-points (T0, T1, and T3). Those were headaches, psychological symptoms (i.e. irritability, apathy, and depression), arterial hypertension and the visual analog scale (as part of EuroQol questionnaire).

From these four features, the headache was the most stable feature (i.e. with the highest stability score) at T0 and T1. In the baseline, all these characteristics were found significant with respect to the CPAP compliance but the latter ($p = 0.079$). In particular, compliant patients were more likely to not having headache (85%, 23 out of 27) nor psychological symptoms (67%, 18 out of 27), having arterial hypertension (74%, 20 out of 27) and worst visual analog scale score (9.15+/-1.02 on mean difference with respect to non-compliant patients). To the best of our knowledge, these features together have not previously been reported as relevant to predict patient compliance with CPAP therapy at either month 0, 1, and 3. In the literature, having morning headache was also found significant in a randomized control trial of OSA patients [42]. On the contrary, psychological factors did not show prediction capability in [43–45] but how patients were challenging difficult situations (active versus passive) [46]. In [47], authors highlighted the positive effect of CPAP treatment on blood pressure in patients with resistant
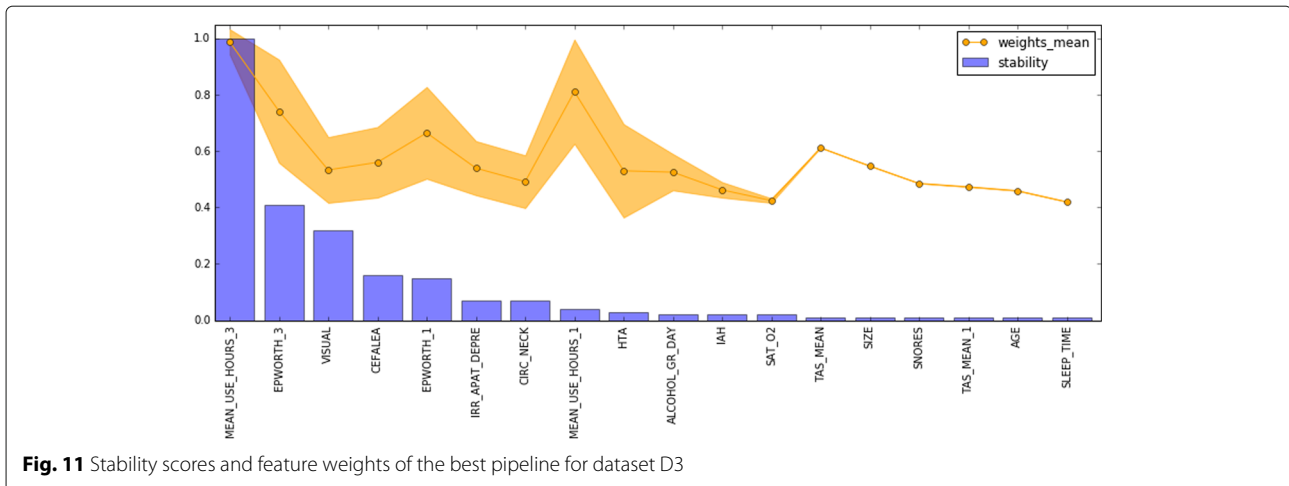

Fig. 10 Stability scores and feature weights of the best pipeline for dataset D1

Rafael-Palou *et al. BMC Medical Informatics and Decision Making* (2018) 18:81

Page 12 of 14

Fig. 11 Stability scores and feature weights of the best pipeline for dataset D3

hypertension. The visual analog scale, used as a generic method for measuring the quality of life, was reported useful to track treatment-induced changes in [48, 49].

Different studies [50–52] have shown an improvement in snoring, gastro-esophageal reflux and oxygen saturation with CPAP treatment. In our case, they were found among the characteristics with the highest stability scores for the best pipelines of month-0 and month-1. However, only oxygen saturation ($p < 0.001$) predicted good compliance with CPAP.

Two of the features collected at months 1 and 3 (i.e. average hours of nightly CPAP use and Epworth) were found as the most important predictive features in these time-points. These features were significant regarding CPAP compliance. Interestingly, from months 1 to 3 the average of nightly hours of use for compliant users increased (from 5.9+/-1.51 to 6.17+/-1.29) while in non-compliant users decreased (from 4.4+/-1.75 to 3.56+/-1.76). On the contrary, the average of Epworth for compliant users decreased from 5.48+/-3.63 to 4.64+/-3.07, while for non-compliant increased from 7.33+/-3.7 to 8.46+/-4.16). Early measurements of the average hours of nightly CPAP use were already reported as predictive of CPAP compliance in [53, 54]. Epworth was also reported as a relevant predictor of compliance in [15, 54–56].

Table 5 Performance comparison between best pipelines for each dataset

Pipelines	Difference (cv_f1)	Statistic	*p*_value
p0 vs p1	0.09 +/- 0.15	-1.71	0.1201
p0 vs p3	0.14 +/- 0.18	-2.70	0.0241
p1 vs p3	0.05 +/- 0.14	-1.0931	0.3027

The limitations of this study can be summarized in the following two points. First, although a common cut-off was selected for the definition of CPAP compliance, changes in this threshold might cause different performances as well as variations in the rank of the feature importance reported in this work, thus further explorations are required in this regard. Second, despite the positive scores obtained by the predictive models at the different time-points, the small number of individuals in the sample makes it highly recommendable to validate the obtained results with new data.

Finally, let us point out that the work proposed in this paper is part of the myOSA project (RTC-2014-3138-1), aimed at developing new ICT tools to support the OSA treatment. Under the umbrella of myOSA, we created an IoT system [57] that remotely monitors the patient's CPAP devices to provide indicators of progress such as early compliance, adherence level, as well as personalized recommendations to empower the patients. Moreover, we are currently investigating how to apply these predictive methods also in the context of the H2020 project CONNECARE (ID: 689802), in which we focus on patient's monitoring with the final goal of providing self-management features to people in needs, such as chronic patients [58].

Conclusions

To the best of our knowledge, this article is the first attempt to analyze and compare the compliance with the CPAP therapy of patients with OSA at different points of the treatment by building classifiers. Three time-points were established to perform the analysis (i.e. before the treatment starts, at month 1, and at month 3). To build and evaluate the classifiers a flexible framework was designed relying on machine learning pipelines. High performances were reached yet after one month of treatment, being the

Rafael-Palou *et al. BMC Medical Informatics and Decision Making* (2018) 18:81

Page 13 of 14

third month when significant differences in performances were achieved with respect to the baseline. Four baseline variables were found relevant for the prediction of compliance with CPAP at each time-point. Two characteristics more, collected during the follow-up, were also highlighted for the prediction of compliance at months 1 and 3. Further tasks are devised to extend the present study, including the collection of new patients and exploring other CPAP compliance cut-offs, in order to validate the findings and reported performances. This work aims to take a step forward towards the creation of new tools to allow early and accurate detection of patients struggling to follow the CPAP treatment and thus enable personalized patient interventions that would lead to improving their quality of life.

Additional file

Additional file 1: Tables with the description of all the variables of the D0, D1 and D3 datasets. Tables with the results of the descriptive analysis of the datasets. (DOCX 944 kb)

Abbreviations

ADO: Oral antidiabetics; AHI: Apnea-hypopnea index; AUC: Area under the curve; CONNECARE: personalized connected care for complex chronic patients; CPAP: Continuous positive pressure; IoT: Internet of things; ICT: Information and communications technology; k-NN: k-nearest neighbor; LASSO: Least absolute shrinkage and selection operator; LR: Logistic regression; MI: Mutual information; ML: Machine learning; myOSA: A telemedicine system for management and treatment of patients with obstructive sleep apnea/hypopnea syndrome NN: Neural network; ODI: Oxygen desaturation index; OSA: Obstructive sleep apnea; RFE: Recursive feature elimination; ROC: receiver operating characteristic; SMOTE: Synthetic minority over-sampling technique; SVM: Support vector machine

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Authors' contributions

All the authors have been involved in drafting the manuscript and revising it. They all contributed in the conception and design of the analysis presented in this paper. XRP performed the analysis of the datasets and worked on the model definition and evaluation. CT gave clinical support in all the phases of the model definition and creation. AS contributed to data collection, labeling and pre-processing. MST and FB supervised the study from the clinical perspective. EV scientifically supervised the study. Finally, all the authors given final approval of the version to be published and agreed to be accountable for all aspects of the work.

Ethics approval and consent to participate

Ethics approval was obtained from the Human Research Ethics Committee of University Hospital of Arnau de Vilanova (Lleida) with the reference number 10/2014, in accordance with the Declaration of Helsinki, and written informed consent to participate was obtained from each patient included in the investigation.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

[1]Eurecat Centre Tecnòlogic de Catalunya, eHealt Unit, Carrrer Bilbao, 72, 08005 Barcelona, Spain. [2]BCN Medtech, Department of Information and Communication Technologies, Universitat Pompeu Fabra, Barcelona, Spain. [3]Institut de Recerca Biomèdica (IRBlleida), Lleida, Spain. [4]CIBERES, Madrid, Spain.

References

1. Smith PL, Hudgel DW, Olson L, Partinen M, Rapoport DM, Rosen CL, Skatrud JB, Waldhorn RE, Westbrook PR, Young T. Indications and standards for use of nasal continuous positive airway pressure (cpap) in sleep apnea syndromes. Am J Respir Crit Care Med. 1994;150(6 I):1738–45.
2. Kribbs NB, Pack AI, Kline LR, Smith PL, Schwartz AR, Schubert NM, Redline S, Henry JN, Getsy JE, Dinges DF. Objective measurement of patterns of nasal cpap use by patients with obstructive sleep apnea. Am Rev Respir Dis. 1993;147(4):887–95.
3. Prosise GL, Berry RB. Oral-nasal continuous positive airway pressure as a treatment for obstructive sleep apnea. CHEST J. 1994;106(1):180–6.
4. Sanders MH, Kern NB, Stiller RA, Strollo PJ, Martin TJ, Atwood CW. Cpap therapy via oronasal mask for obstructive sleep apnea. CHEST J. 1994;106(3):774–9.
5. Sullivan C, Berthon-Jones M, Issa F, Eves L. Reversal of obstructive sleep apnoea by continuous positive airway pressure applied through the nares. Lancet. 1981;317(8225):862–5.
6. Engleman HM, Martin SE, Kingshott RN, Mackay TW, Deary IJ, Douglas NJ. Randomised placebo controlled trial of daytime function after continuous positive airway pressure (cpap) therapy for the sleep apnoea/hypopnoea syndrome. Thorax. 1998;53(5):341–5. https://doi.org/10.1136/thx.53.5.341. http://thorax.bmj.com/content/53/5/341.full.pdf.
7. Marin JM, Carrizo SJ, Vicente E, Agusti AG. Long-term cardiovascular outcomes in men with obstructive sleep apnoea-hypopnoea with or without treatment with continuous positive airway pressure: an observational study. Lancet. 2005;365(9464):1046–53.
8. Hussain SF, Irfan M, Waheed Z, Alam N, Mansoor S, Islam M. Compliance with continuous positive airway pressure (cpap) therapy for obstructive sleep apnea among privately paying patients-a cross sectional study. BMC Pulm Med. 2014;14(1):188.
9. Campos-Rodriguez F, Martinez-Alonso M, Sanchez-de-la-Torre M, Barbe F, Network SS, et al. Long-term adherence to continuous positive airway pressure therapy in non-sleepy sleep apnea patients. Sleep Med. 2016;17:1–6.
10. Weaver TE, Chasens ER. Continuous positive airway pressure treatment for sleep apnea in older adults. Sleep Med Rev. 2007;11(2):99–111.
11. McArdle N, Devereux G, Heidarnejad H, Engleman HM, Mackay TW, Douglas NJ. Long-term use of cpap therapy for sleep apnea/hypopnea syndrome. Am J Respir Crit Care Med. 1999;159(4):1108–14.
12. Kohler M, Smith D, Tippett V, Stradling JR. Predictors of long-term compliance with continuous positive airway pressure. Thorax. 2010;65(9):829–32.
13. Krieger J, Kurtz D, Petiau C, Sforza E, Trautmann D. Long-term compliance with cpap therapy in obstructive sleep apnea patients and in snorers. Sleep. 1996;19(suppl_9):136–43.

Rafael-Palou *et al. BMC Medical Informatics and Decision Making*   (2018) 18:81

Page 14 of 14

14. Chen Y-F, Hang L-W, Huang C-S, Liang S-J, Chung W-S. Polysomnographic predictors of persistent continuous positive airway pressure adherence in patients with moderate and severe obstructive sleep apnea. Kaohsiung J Med Sci. 2015;31(2):83–9.

15. Budhiraja R, Parthasarathy S, Drake CL, Roth T, Sharief I, Budhiraja P, Saunders V, Hudgel DW. Early cpap use identifies subsequent adherence to cpap therapy. Sleep. 2007;30(3):320–4.

16. Çinar M, Engin M, Engin EZ, Ateşçi YZ. Early prostate cancer diagnosis by using artificial neural networks and support vector machines. Expert Syst Appl. 2009;36(3):6357–61.

17. Hassanien AE, Moftah HM, Azar AT, Shoman M. Mri breast cancer diagnosis hybrid approach using adaptive ant-based segmentation and multilayer perceptron neural networks classifier. Appl Soft Comput. 2014;14:62–71.

18. Lee SK, Kang B-Y, Kim H-G, Son Y-J. Predictors of medication adherence in elderly patients with chronic diseases using support vector machine models. Healthcare Inform Res. 2013;19(1):33–41.

19. Bourdes V, Ferrieres J, Amar J, Amelineau E, Bonnevay S, Berlion M, Danchin N. Prediction of persistence of combined evidence-based cardiovascular medications in patients with acute coronary syndrome after hospital discharge using neural networks. Med Biol Eng Comput. 2011;49(8):947–55.

20. Cortes C, Vapnik V. Support-vector networks. Mach Learn. 1995;20(3): 273–97.

21. Bishop C, Bishop CM, et al. Neural networks for pattern recognition: Oxford University Press; 1995.

22. Auria L, Moro RA. Support vector machines (SVM) as a technique for solvency analysis; 2008. DIW Berlin, German Institute for Economic Research.

23. Cox DR. The regression analysis of binary sequences. J R Stat Soc Ser B (Methodol). 1958:215–242.

24. Rokach L, Maimon O. Data mining with decision trees: theory and applications. Data Mining with Decision Trees: Theory and Applications. Edited by Lior Rokach and Oded Maimon. Published by World Scientific Publishing Co. Pte. Ltd.,# 9789814590082, 2014. 2014.

25. Ross BC. Mutual information between discrete and continuous data sets. PloS ONE. 2014;9(2):87357.

26. Inza I, Calvo B, Armañanzas R, Bengoetxea E, Larrañaga P, Lozano JA. Machine learning: an indispensable tool in bioinformatics. Springer; 2010. p. 25–48.

27. Bermingham ML, Pong-Wong R, Spiliopoulou A, Hayward C, Rudan I, Campbell H, Wright AF, Wilson JF, Agakov F, Navarro P, et al. Application of high-dimensional feature selection: evaluation for genomic prediction in man. Sci Rep. 2015;5:10312.

28. Guyon I, Elisseeff A. An introduction to variable and feature selection. J Mach Learn Res. 2003;3(Mar):1157–82.

29. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science. 1999;286(5439):531–7.

30. Tibshirani R. Regression shrinkage and selection via the lasso. J R Stat Soc Ser B (Methodol). 1996:267–88.

31. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. Smote: synthetic minority over-sampling technique. J Artif Intell Res. 2002;16:321–57.

32. Altman NS. An introduction to kernel and nearest-neighbor nonparametric regression. Am Stat. 1992;46(3):175–85.

33. Ho TK. Random decision forests. Document Analysis and Recognition, 1995. In: Proceedings of the Third International Conference On. vol 1. IEEE; 1995. p. 278–82.

34. Friedman J, Hastie T, Tibshirani R. The elements of statistical learning, . Vol 1, Issue 10. New York: Springer series in statistics; 2001.

35. Cawley GC, Talbot NL. On over-fitting in model selection and subsequent selection bias in performance evaluation. J Mach Learn Res. 2010;11(Jul): 2079–107.

36. Meinshausen N, Bühlmann P. Stability selection. J R Stat Soc Ser B (Stat Methodol). 2010;72(4):417–73.

37. Raudys SJ, Jain AK, et al. Small sample size effects in statistical pattern recognition: Recommendations for practitioners. IEEE Trans Pattern Anal Mach Intell. 1991;13(3):252–64.

38. Cawley GC, Talbot NL. On over-fitting in model selection and subsequent selection bias in performance evaluation. J Mach Learn Res. 2010;11(Jul): 2079–107.

39. Weaver TE, Kribbs NB, Pack AI, Kline LR, Chugh DK, Maislin G, Smith PL, Schwartz AR, Schubert NM, Gillen KA, et al. Night-to-night variability in cpap use over the first three months of treatment. Sleep. 1997;20(4): 278–83.

40. Schölkopf B, Smola AJ. Learning with kernels: support vector machines, regularization, optimization, and beyond. MIT; 2002.

41. Weber R, Schek H-J, Blott S. A quantitative analysis and performance study for similarity-search methods in high-dimensional spaces. In: Proc. of the VLDB conference, New York, 1998; 1998. p. 194–205.

42. Kristiansen HA, Kværner KJ, Akre H, Øverland B, Sandvik L, Russell MB. Sleep apnoea headache in the general population. Cephalalgia. 2012;32(6):451–8.

43. Stepnowsky C, Marler MR, Ancoli-Israel S. Determinants of nasal cpap compliance. Sleep Med. 2002;3(3):239–47.

44. Aloia M, Arnedt J, Stepnowsky C, Hecht J, Borrelli B. Predicting treatment adherence in obstructive sleep apnea using principles of behavior change. J Clin Sleep Med JCSM Off Publ Am Acad Sleep Med. 2005;1(4):346–53.

45. Lazarus RS, Folkman S. Coping and adaptation. The handbook of behavioral medicine. Vol 282325. New York: Guilford; 1984.

46. Stepnowsky Jr CJ, Bardwell WA, Moore PJ, Ancoli-Israel S, Dimsdale JE. Psychologic correlates of compliance with continuous positive airway pressure. Sleep. 2002;25(7):758–62.

47. Martínez-García MA, Capote F, Campos-Rodríguez F, Lloberes P, de Atauri MJD, Somoza M, Masa JF, González M, Sacristán L, Barbé F, et al. Effect of cpap on blood pressure in patients with obstructive sleep apnea and resistant hypertension: the hiparco randomized clinical trial. Jama. 2013;310(22):2407–15.

48. Schmidlin M, Fritsch K, Matthews F, Thurnheer R, Senn O, Bloch KE. Utility indices in patients with the obstructive sleep apnea syndrome. Respiration. 2010;79(3):200–8.

49. Chakravorty I, Cayton R, Szczepura A. Health utilities in evaluating intervention in the sleep apnoea/hypopnoea syndrome. Eur Respir J. 2002;20(5):1233–8.

50. Tamanna S, Campbell D, Warren R, Ullah MI. J Clin Sleep Med JCSM Off Publ Am Acad Sleep Med. 2016;12(9):1257.

51. Li D, Liu D, Wang X, He D. Self-reported habitual snoring and risk of cardiovascular disease and all-cause mortality. Atherosclerosis. 2014;235(1):189–95.

52. Campos-Rodriguez F, Pena-Grinan N, Reyes-Nunez N, De la Cruz-Moron I, Perez-Ronchel J, De la Vega-Gallardo F, Fernandez-Palacin A. Mortality in obstructive sleep apnea-hypopnea patients treated with positive airway pressure. CHEST J. 2005;128(2):624–33.

53. Chai-Coetzer CL, Luo Y-M, Antic NA, Zhang X-L, Chen B-Y, He Q-Y, Heeley E, Huang S-G, Anderson C, Zhong N-S, et al. Predictors of long-term adherence to continuous positive airway pressure therapy in patients with obstructive sleep apnea and cardiovascular disease in the save study. Sleep. 2013;36(12):1929–37.

54. Ghosh D, Allgar V, Elliott M. Identifying poor compliance with cpap in obstructive sleep apnoea: a simple prediction equation using data after a two week trial. Respir Med. 2013;107(6):936–42.

55. Popescu G, Latham M, Allgar V, Elliott M. Continuous positive airway pressure for sleep apnoea/hypopnoea syndrome: usefulness of a 2 week trial to identify factors associated with long term use. Thorax. 2001;56(9): 727–33.

56. Weaver TE, Maislin G, Dinges DF, Bloxham T, George CF, Greenberg H, Kader G, Mahowald M, Younger J, Pack AI. Relationship between hours of cpap use and achieving normal levels of sleepiness and daily functioning. Sleep. 2007;30(6):711–9.

57. Rafael-Palou X, Vargiu E, Turino C, Steblin A, Sánchez-de-la-Torre M, Barbé F. Towards an intelligent monitoring system for patients with obstructive sleep apnea. ICST Trans Ambient Syst. 2017;4:1.

58. Vargiu E, Fernández JM, Miralles F, Nakar S, Weijers V, Meetsma H, Mariani S, Mamei M, Zambonelli F, Michel F, Matthes F, Kelly J, Eaglesham J, Kaye R. Patient empowerment and case management in connecare. In: International Conference on Integrated Care, Utrecht, May 23-25, 2018; 2018.