

# An Evaluation of Indexes as Support Tools in the Diagnosis of Sleep Apnea

ABRAHAM OTERO,<sup>1</sup> PAULO FÉLIX,<sup>2</sup> JESÚS PRESEDO,<sup>2</sup> and CARLOS ZAMARRÓN<sup>3</sup>

<sup>1</sup>Department of Information and Communications Systems Engineering, University San Pablo CEU, Boadilla del Monte, Km 5,300, Urbanización Montepríncipe, 28668 Madrid, Spain; <sup>2</sup>Centro de Investigación en TecnoloXías da Información (CITIUS), 15782 Santiago, Spain; and <sup>3</sup>Division of Respiratory Medicine, University Hospital Complex of Santiago de Compostela, 15782 Santiago, Spain

(Received 12 August 2011; accepted 21 February 2012; published online 29 February 2012)

Associate Editor Leonidas D. Iasemidis oversaw the review of this article.

**Abstract**—This article evaluates several indexes as support tools to diagnose patients with Sleep Apnea-Hypopnea Syndrome (SAHS). Some of these indexes, such as the Apnea-Hypopnea Index, have been standardized and studied in depth in the literature. Other indexes are used extensively in the reports that commercial polysomnographs generate. However, they have not been studied in detail and clinicians have no standardized guidelines for interpreting them. Examples are the mean and maximum duration of apneas and hypopneas. Finally, several novel indexes proposed by the authors are also evaluated. To evaluate the indexes, we have used a database of 274 patients who have undergone a polysomnographic test. Several feature selection techniques were used to assess the capability of each index to discriminate between healthy and SAHS patients. The capability of the indexes for diagnosing the patients was analyzed by using decision trees which were trained using each index individually, and all the indexes together. Our results suggest that some indexes which are often present in the reports of commercial polysomnographs provide little or no information. On the other hand, other indexes that are usually not considered have a great capability to discern between SAHS and control patients.

**Keywords**—Sleep Apnea-Hypopnea, Apnea-Hypopnea Index, Polysomnography.

## INTRODUCTION

Sleep Apnea-Hypopnea Syndrome (SAHS) is a sleep-breathing disorder characterized by recurrent episodes of upper airway narrowing or collapsing

during sleep that produce a total or partial cessation of the patient's respiratory airflow (RA).<sup>12</sup> When the cessations are total they are called apneas; when they are partial, they are called hypopneas. Apneas and hypopneas are usually accompanied by hypoxemia, with a drop in blood oxyhemoglobin saturation (SpO<sub>2</sub>).<sup>31</sup> This disorder is estimated to affect 4% of the adult male population and 2% of the adult female population.<sup>34</sup>

The overall result of SAHS is a disruption in the patient's sleep architecture and a decrease in its refreshing effects. Consequently, patients often suffer from daytime drowsiness and cognitive deficits which increase the risks of accidents in the workplace and when driving vehicles.<sup>10</sup> They may also suffer from depression, anxiety, excessive irritability and various sexual dysfunctions. Treatment of SAHS depends on the severity of the illness: in lighter cases, changes in the patient's behaviour (e.g. losing weight, avoiding alcoholic beverages and avoiding sleeping positions likely to trigger apneas, among others) may be sufficient; in the most serious cases, it may be necessary to resort to surgery and, more frequently, to therapy with Continuous Positive Airway Pressure (CPAP). A CPAP device applies pressure (constant in the case of older models, variable following inhalations and exhalations of the patient in newer models) on the airway by means of a nasal mask while the patient is asleep. This pressure prevents the collapse of the upper airway, thus avoiding apneas. Once a patient has started to use CPAP therapy, he/she may often have to continue using it for the rest of his/her life.

The gold standard test for the diagnosis of SAHS is the polysomnography, a test performed in a Hospital Sleep Unit that consists of the recording of a wide

---

Address correspondence to Abraham Otero, Department of Information and Communications Systems Engineering, University San Pablo CEU, Boadilla del Monte, Km 5,300, Urbanización Montepríncipe, 28668 Madrid, Spain. Electronic mail: abraham.otero@gmail.com

range of physiological parameters while the patient is asleep.<sup>15</sup> Commercial polysomnographs usually generate a report summarizing the test. This report together with a visual inspection of the polysomnogram are the main tools that the clinician uses to diagnose the patient.

The most relevant piece of information contained in these reports is very likely the Apnea-Hypopnea Index (AHI), i.e., the number of total or partial interruptions of the RA that the patient experiences per hour of sleep. This index has been the target of comprehensive clinical studies, and there are detailed clinical guidelines that are used to interpret it.<sup>6</sup> The AHI is considered by clinicians to determine whether the patient suffers from SAHS, as well as to evaluate the severity of his/her condition. In accordance with the criteria of the American Academy of Sleep Medicine, a patient may be diagnosed with SAHS if he/she has five or more apneas during each hour of sleep throughout the entire night.<sup>3</sup> However, clinicians do not usually make decisions about the patient's condition on the basis of the AHI alone, but they also take into account the patients' symptomatology and anthropometric information, laboratory analysis, and descriptions of the patient's partner about his/her sleep (snoring, presence of breathing pauses, etc.), among others.<sup>10</sup>

The main weakness of the AHI is that it is based on a simple count of events, not on a characterization of those events. A respiratory pause of 50 s and a respiratory pause of 10 s contribute equally to the AHI. Therefore, under certain circumstances this index can underestimate the severity of the patient's condition.<sup>28</sup> Hence the interest in having available other indexes that could complement the information provided by the AHI.

The reports generated by commercial polysomnographs contain other indexes besides the AHI. These are mainly mean and maximum values of certain descriptors of the pathological events that have occurred during the night; e.g., the mean and maximum duration of apneas, hypopneas and desaturations, the mean of the minimum SpO<sub>2</sub> value reached in each desaturation, *et cetera*. Many authors have studied how to provide support for the diagnosis of SAHS by analyzing patients' anthropometric information, data extracted from questionnaires, and laboratory analysis.<sup>7,19,35,36</sup> However, to the best of our knowledge, comprehensive studies on how to take advantage of many of the indexes that are presented to clinicians have not been carried out. Therefore, there are no clear guidelines on how to interpret them. This severely hampers their utility as support tools to the diagnosis of SAHS.

This paper aims to evaluate the usefulness of some of these indexes in the diagnosis of SAHS. The evaluation will also include indexes not usually considered in the

diagnosis of SAHS, but which the authors believe may be useful. Between the former are indexes based on a similar concept as the AHI, but they incorporate information about the duration of the pathological events and their severity. Examples of these indexes are the percentage of time that the patient has been in apnea, in hypopnea or in desaturation.

We have generated the indexes targeted in this study for 274 patients who underwent polysomnographic testing. No distinction between obstructive and central sleep apnea patients was made during the compilation of the database. Each of those patients was diagnosed by a pulmonologist. Several feature selection techniques were used to evaluate the capability of each index to discriminate between healthy and SAHS patients. The capability of the indexes for predicting the class was also analyzed by using decision trees.

"Methods" section describes the patient database used in our study and presents the materials and methods of the analysis we have performed. "Results" section presents the results obtained, and "Discussion" section discusses them. Finally, a series of conclusions on the paper are given.

## METHODS

### Data

The database used in this paper consists of a total of 274 polysomnographic recordings arising from 274 patients who were subjected to a sleep study in the Sleep Unit of the University Hospital Complex of Santiago de Compostela. Access to the data was approved by the hospital's Clinical Research Ethics Committee after evaluating the research protocol. The recordings of all patients who underwent polysomnographic testing in the Sleep Unit during the second half of 2009 and the first half of 2010 were considered for inclusion in the database. To obtain a more homogeneous data base, patients with other sleep disorders besides SAHS which may cause excessive daytime sleepiness, such as insomnia, narcolepsy or restless legs syndrome, were excluded from the database. Patients with liver disease, kidney failure, psychiatric disorders, vascular disease, diabetes mellitus or thyroid gland abnormalities, drug addiction, alcoholism (>60 g alcohol/day), or patients who used hypnotics or sedative drugs regularly also were excluded. All the remaining patients were included in the database. No distinction between obstructive and central sleep apnea patients was made.

Recordings were performed with the commercial polysomnograph SOMNOscreen<sup>TM</sup>, built by SOMNOmedics GmbH. Oronasal flow was registered by nasal prongs. Patients also were wearing chest and

abdomen piezoelectric belts to register thoracic and abdominal movements. Body position, oximetry, electrocardiogram and snoring also were recorded. The recording started a few minutes after turning off the room lights when the patient had fallen asleep (5–20 min). The study finished when the patient awoke. 23 patients did the study during a brief nap in the afternoon; their recordings have an average duration of approximately 3 h. The remaining patients underwent the test during the night; their recordings have an average duration of approximately 8 h. 224 patients were males and 50 were females. Their average age was  $53.2 \pm 12.7$  (mean  $\pm$  std) years, with a minimum age of 23 and maximum of 88. Their average weight was  $89.9 \pm 16.4$  kg, with a average body mass index (BMI) of  $32.4 \pm 6.1$  kg/m<sup>2</sup>.

A pulmonologist with over 20 years of experience working in the Sleep Unit diagnosed the patients in our database. The diagnostic criteria used were:

- If  $AHI < 5$  the patient does not have SAHS.
- If  $5 \leq AHI \leq 10$  the patient has SAHS if (1) the patient is obese and snored during the night; or (2) if the patient is obese and presented significant drops in SpO<sub>2</sub> during the apneas/hypopneas; or (3) if the patient snored during the night and presented significant drops in SpO<sub>2</sub> during the apneas/hypopneas; or (4) if the patient is obese and he/she reported excessive daytime sleepiness, then the patient has SAHS. Otherwise, the patient does not have SAHS.
- If  $AHI > 10$  the patient has SAHS.

Obesity was defined as  $BMI > 30$  kg/m<sup>2</sup>. Drops in SpO<sub>2</sub> were considered significant if the average fall in SpO<sub>2</sub> was  $> 5\%$ . Snoring was considered to be present during the polysomnography if the patient snored during at least 3 min/h sleep. Using these criteria, 207 patients were diagnosed with SAHS; the remaining

67 did not suffer from SAHS. More information about the database used in this paper can be found in Appendix.

### Indexes Generation

To generate the indexes targeted in this study for the 274 patients, their polysomnographic recordings were analyzed using algorithms previously developed by the authors for this purpose (see Fig. 1). In the literature there are similar proposals that generate the same information.<sup>2,4,5,22,23,32,33</sup> Our algorithms are capable of identifying apneas, hypopneas, desaturations, thoracic and abdominal movement limitations and snoring in the polysomnogram.<sup>25</sup> For the study presented in this paper, they were only used to identify apneas, hypopneas and desaturations. In addition to identifying each individual event, a highly robust association between the hypoventilations—apneas or hypopneas—and the desaturations they cause was performed with the algorithms presented in Otero and Félix.<sup>24</sup>

In the identification of apneas, hypopneas and desaturations our algorithms present a false positive rate of 2.6, 6.0, and 1.2% respectively.<sup>25</sup> The false negative rate is 0.9, 3.4, and 1.1% respectively.<sup>25</sup> When performing the association between apneas/hypopneas and the desaturations they cause, the rate of false positives in the identification of compound events falls to 0.86%. This low percentage of false positives is achieved thanks to the integration of information arising from two signals: respiratory airflow and SpO<sub>2</sub>.<sup>24</sup>

Using these algorithms, 46,505 associations between a respiratory airflow limitation—either an apnea or a hypopnea—and a desaturation were generated. For each of them, the following descriptors were generated (see Fig. 2):

- DurFlux: duration of the respiratory airflow limitation measured in seconds.

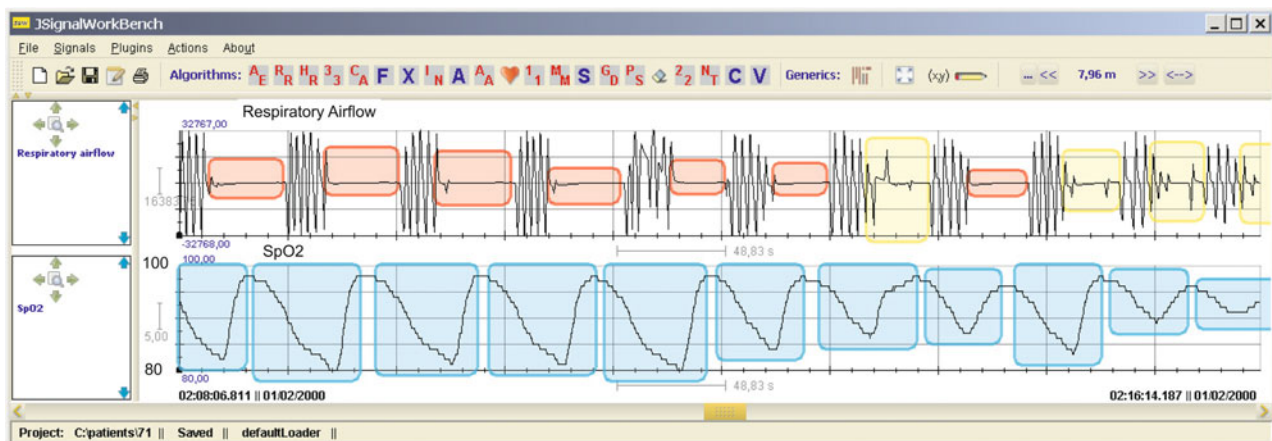
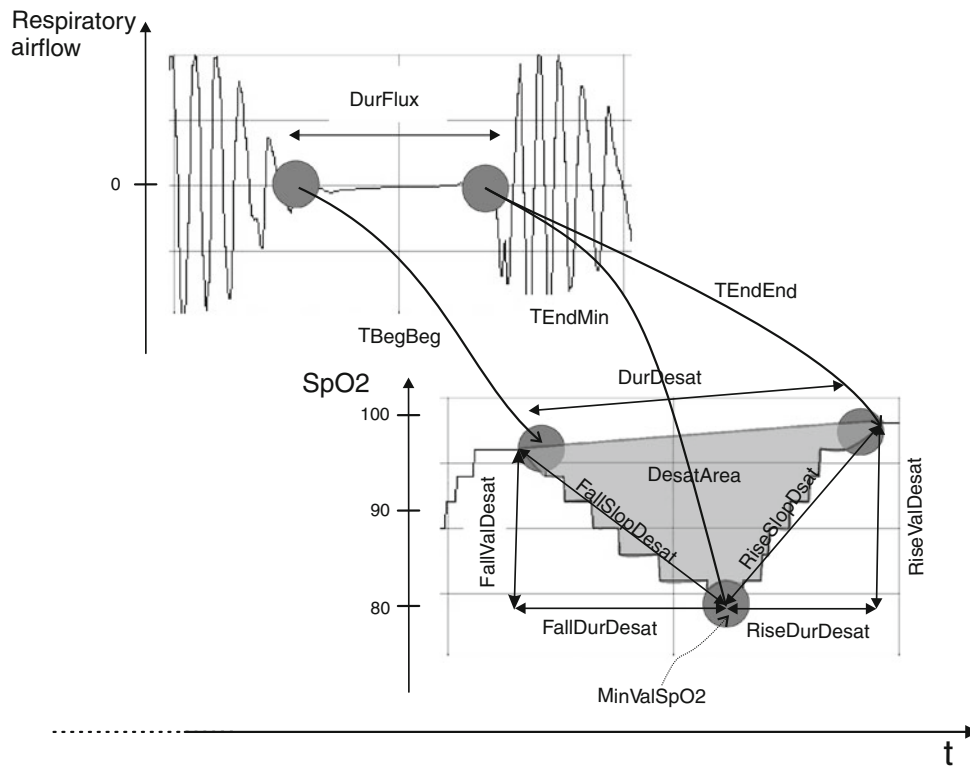


FIGURE 1. Tool which implements the algorithms used to identify apneas, hypopneas and desaturations.



**FIGURE 2.** Set of descriptors generated for each association between an apnea and a desaturation.

- DurDesat: duration of the desaturation measured in seconds.
- MinValSpO2: minimum value of the SpO2 during the desaturation.
- MeanValSpO2: mean value of the SpO2 during the desaturation.
- FallDurDesat: time elapsed from the beginning of the desaturation until the minimum value of the SpO2 is reached—fall section of the desaturation.
- FallValDesat: drop in SpO2 during the fall section of the desaturation.
- FallSlopDesat: slope of the fall section of the desaturation;  $(-FallValDesat)/FallDurDesat$ .
- RiseDurDesat: time elapsed from the point where the minimum value of the desaturation is reached until the desaturation ends (rise section of the desaturation).
- RiseValDesat: increase in SpO2 during the rise section of the desaturation.
- RiseSlopDesat: slope of the rise section of the desaturation;  $RiseValDesat/RiseDurDesat$ .
- DesatArea: area between the straight line joining the starting and ending points of a desaturation and the SpO2 signal (gray area in Fig. 2). This area increases with the duration of the desaturation, and with the magnitude of the drop in SpO2 from the basal value of this

parameter. Therefore, this index attempts to quantify the severity of the desaturations merging information from its duration and its fall in SpO2 in a single parameter. Its units are (% saturation)  $\times$  second.

- TBegBeg: time elapsed between the beginning of the respiratory airflow limitation and the beginning of the desaturation measured in seconds.
- TEndMin: time elapsed between the end of the respiratory airflow limitation and the instant at which the minimum value of the desaturation is reached measured in seconds.
- TEndEnd: time elapsed between the end of the respiratory airflow limitation and the end of the desaturation measured in seconds.

All these descriptors were generated for each of the 46,505 associations between pathological events. Then, from the events belonging to each patient, we calculated the average value of each descriptor for that patient. The average values were calculated independently for associations between apneas and desaturations, and for associations between hypopneas and desaturations. Previous work by the authors suggests that it is useful to distinguish between events caused by apneas and by hypopneas, as they seem to have different capabilities for discriminating between control

and SAHS patients.<sup>26</sup> In this paper we will use the nomenclature of the previous list to refer to measures taken over an event that was triggered by an apnea, while for events triggered by hypopneas we add an “H” at the end of the names of the list. For example, DurDesat is the duration of a desaturation that has been caused by an apnea, while DurDesatH is the duration of a desaturation that has been caused by a hypopnea.

We also calculated for each patient the maximum duration of the apneas (DurFluxM) and hypopneas (DurFluxHM) he/she experienced; and the maximum duration of the desaturations caused by apneas (DurDesatM), and by hypopneas (DurDesatHM). These maximum values were included in the analysis because they often appear in the reports of commercial polysomnographs. Finally, we also generated the following pieces of information for each patient:

- Standard AHI.
- MeanValSpO2: mean value of the SpO2 throughout the recording.
- BasalSpO2: basal value of the SpO2.
- DeltaSpO2: difference between the SpO2 basal value and the mean value of the SpO2 throughout the recording ( $\text{BasalSpO2} - \text{MeanValSpO2}$ ).
- ApPercent: percentage of sleep time that the patient has been in apnea.
- HypopPercent: percentage of sleep time that the patient has been in hypopnea.
- ApHypopPercent:  $\text{ApPercent} + \text{HypopPercent}$ .
- DesatPercent: percentage of sleep time that the patient has been in hypoxia.
- AHDI (Apnea-Hypopnea-Desaturation Index):  $\text{ApPercent} + \text{HypopPercent} + \text{DesatPercent}$ .
- AreaSpO2: area between a horizontal line with a value equal to BasalSpO2 and the oxygen saturation curve, normalized by the number of hours of sleep.

To calculate ApPercent, HypopPercent, ApHypopPercent, DesatPercent, and AHDI all the apneas, hypopneas and/or desaturations identified in the patient’s polysomnogram were considered, and not only those apneas and/or hypopneas that were successfully associated with their corresponding desaturation (i.e., no events were excluded in the calculation of these indexes). These indexes attempt to characterize the quantity and the duration of the pathological events that the patient has experienced throughout the night. Therefore, they must consider all events. The indexes that describe the associations between respiratory airflow limitations and desaturations try to characterize the average features of these associations for each patient. It is not necessary to consider all events when calculating the average features. That is the reason why the algorithms

we have used for event association have a very high specificity (0.83% of false positives), though they discard a larger number of events on average: only 88% of the airflow obstructions are successfully associated with the desaturation they have caused.<sup>24</sup>

The SpO2 basal value was calculated as follows. All SpO2 samples recorded during the patient’s sleep are sorted. The 90% of the lowest values are disregarded on the grounds that they may contain desaturations. The 5% of the highest values are disregarded on the ground that they may contain hyperventilation periods. The patient’s basal value is calculated as the average of the remaining 5% of the samples.

Each of the 274 patients was represented by a feature vector containing the 42 indexes described in this section. For each of the 42 indexes we have 67 measures taken from non-SAHS patients and 207 measures taken from SAHS patients. Therefore, the task of evaluating how relevant each of these features is in the diagnosis of SAHS can be seen as a feature evaluation problem: we want to evaluate the capability of each feature—index—to discriminate between the two classes.

### *Index Evaluation*

To evaluate the capability of each index to discriminate between healthy and SAHS patients we shall use several feature selection techniques. We shall also train a classifier to recognize SAHS and control patients using each of the indexes individually, and all the indexes together. Each classifier’s performance when applied to our patient database will be measured.

### *Feature Selection*

We shall evaluate each index using several feature evaluation techniques that are capable of ranking the features according to the information they provide: Information Gain, Gain Ratio, Chi-square feature evaluation, Relief feature evaluation, and a support vector machine-based (SVM) technique.

Information Gain and Gain Ratio try to measure the information obtained when making a decision based on a given feature.<sup>30</sup> Gain Ratio is based on Information Gain, but it applies a correction to penalize decisions with a high branching factor. Chi-square feature evaluation ranks features by computing the value of the Chi-squared statistic with respect to the class.<sup>20</sup> These techniques can only be applied to discrete features. All the indexes that we are evaluating are continuous. Therefore, to apply them the indexes were discretized with the algorithm presented in Fayyad and Irani.<sup>9</sup>

Relief evaluates the features by repeatedly sampling an instance and considering the value of the given feature for the nearest instance of the same class and of a different class.<sup>16</sup> Finally, a SVM-based technique was used.<sup>14</sup> An SVM classifier was trained with the data, and the features were ranked by the square of the weight assigned to each feature by the SVM.

To obtain some type of aggregate score of the results of all these feature evaluation techniques we shall proceed as follows: the index that ranks highest according to a certain technique is assigned 10 points; the second highest ranked index is assigned 9 points ... and the 10th highest ranked index is assigned 1 point. The indexes that are not among the top 10 are assigned 0 points. For each index, the points obtained with the five techniques are added. This sum will be an ad hoc aggregate score of each index.

Finally we shall also use several feature selection techniques: Correlation-based Feature Selection (CFS), Consistency-based Subset Evaluation and a wrapper method. CFS evaluates a subset of features by considering the individual predictive ability of each feature along with the degree of redundancy among the features.<sup>13</sup> Consistency-based Subset Evaluation evaluates a subset of features by the level of consistency in the class values when the instances are projected onto the subset of features.<sup>1</sup> The wrapper method uses a classifier to evaluate feature sets.<sup>17</sup> We used a C 4.5 pruned tree as the classifier for the wrapper method. In all cases the search for subsets of features was performed using greedy forward hill climbing search through the space of feature subsets.

#### *Classification Performance*

We shall evaluate the capability of the indexes that have been selected in the feature selection process to classify patients as SAHS or control. We shall consider that an index has been selected if (1) the index is one of the 10 which have ranked higher according to our ad hoc aggregate score for the feature evaluation techniques; or if (2) the index was selected by any of the

feature selection techniques. To evaluate the prediction capability of each index as an isolated piece of information, we shall measure the performance of a C 4.5 pruned decision tree<sup>30</sup> built using only that index in a 10 fold validation over our patient database. Accuracy, sensitivity, specificity, likelihood ratio positive (LR+) and likelihood ratio negative (LR-) will be calculated for the decision tree.

In the case of the AHI, besides training a C 4.5 pruned decision tree, we also shall evaluate the performance of using the criteria “SAHS if AHI > 5” and “SAHS if AHI > 10” over our patient database.

Finally, we shall also evaluate the performance of decision trees built with (1) all the indexes, (2) all the indexes except AHI and AHDI, and (3) only those indexes that are average and maximum values of the descriptors of apneas, hypopneas and desaturations; i.e. the latter decision tree does not use AHI, AHDI, ApHypopPercent, DesatPercent, HypopPercent, ApPercent, MeanValSpO2, BasalSpO2, DeltaSpO2, and AreaSpO2. Therefore, this decision tree has information about the characteristics of the pathological events experienced by the patient, but not about the number of events that have occurred.

## RESULTS

Table 1 shows the 10 indexes ranked highest by each feature evaluation technique. The indexes are listed from higher ranking to lower ranking. The last column shows the aggregate score of the 10 features that have achieved the highest aggregate scores.

Table 2 shows the indexes selected by the three feature selection techniques: CFS, Consistency-based Subset Evaluation and a wrapper method. The last two techniques only selected one index: AHDI.

Table 3 shows the accuracy, sensitivity, specificity, likelihood ratio positive (LR+) and likelihood ratio negative (LR-) of the decision trees. The first column of the table indicates with which index/indexes the decision tree was built. The “All indexes” decision tree

**TABLE 1. Rankings obtained by several feature evaluation techniques (top 10 ranked features).**

SVM feature evaluator	Relief	Chi-square	Info gain	Gain ratio	Overall (aggregate score)
AHDI	AHDI	AHDI	AHDI	AHDI	AHDI (50)
ApHypopPercent	DesatPercent	AHI	AHI	AHI	ApHypopPercent (41)
DesatPercent	ApHypopPercent	ApHypopPercent	ApHypopPercent	ApHypopPercent	AHI (40)
HypopPercent	AHI	DesatPercent	DesatPercent	HypopPercent	DesatPercent (37)
AHI	HypopPercent	HypopPercent	HypopPercent	DesatPercent	HypopPercent (32)
DurFluxHM	AreaSpO2	RiseValDesatH	ApPercent	FallValDesatH	DeltaSpO2 (14)
DeltaSpO2	DeltaSpO2	ApPercent	RiseValDesatH	RiseValDesatH	ApPercent (13)
DurDesat	ApPercent	DeltaSpO2	DeltaSpO2	RiseDurDesatH	RiseValDesatH (13)
AreaSpO2	MeanValSpO2	AreaSpO2	AreaSpO2	AreaSpO2	AreaSpO2 (11)
DurDesatHM	DurFluxHM	FallValDesatH	FallValDesatH	ApPercent	FallValDesatH (7)

**TABLE 2. Descriptors selected by each of the feature selection techniques.**

CFS	Consistency subset	Wrapper (C 4.5)
RiseValDesatH DurFluxHM ApHypopPercnt AHI AHDH	AHDI	AHDI

**TABLE 3. Performance of the decision trees.**

Index	Accuracy	Sensitivity	Specificity	LR+	LR-
All indexes	0.985	0.995	0.955	22.2	0.00506
All-{AHI, AHDI}	0.971	0.981	0.940	16.4	0.0206
Descriptors	0.836	0.918	0.582	2.20	0.141
AHDI	0.985	0.995	0.955	22.2	0.00506
AHI	0.960	0.976	0.910	10.9	0.0265
AHI $\geq 5$	0.942	1	0.761	4.19	0
AHI $\geq 10$	0.912	0.884	1	$\infty$	0.116
ApHypopPercnt	0.942	0.928	0.985	62.1	0.0736
DesatPercnt	0.909	0.957	0.761	4.00	0.0571
HypopPercnt	0.916	0.990	0.687	3.16	0.0141
ApPercnt	0.865	0.923	0.687	2.94	0.113
DeltaSpO2	0.821	0.903	0.567	2.09	0.170
AreaSpO2	0.821	0.932	0.478	1.78	0.142
DurFluxHM	0.901	0.952	0.746	3.75	0.0647
RiseValueDesatH	0.839	0.947	0.507	1.92	0.105
FallValueDesatH	0.810	0.865	0.642	2.41	0.211
DurFluxM*	0.755	1	0	1	-
DurDesatM*	0.755	1	0	1	-
DurFlux*	0.755	1	0	1	-
DurFluxH*	0.755	1	0	1	-
DurDesat*	0.755	1	0	1	-
DurDesatHM*	0.785	1	0.119	1.13	0
DurDesatH*	0.770	0.951	0.209	1.20	0.231
MinValDesat*	0.770	0.918	0.313	1.34	0.262
MinValDesatH*	0.810	0.966	0.329	1.44	0.103
MeanValSpO2*	0.755	1	0	1	-
BasalSpO2*	0.755	1	0	1	-

\*These indexes were not selected in the feature selection process but are included here for being often present in the reports of commercial polysomnographs.

was built using all the indexes. However, the resulting decision tree selected only one index: AHDI. The “All-{AHI, AHDI}” decision tree was built using all the indexes but AHI and AHDI. The resulting decision tree has seven nodes and four leaves. It uses only ApHypopPercnt and DesatPercnt, which combined provide the same information as AHDI. The “Descriptors” decision tree was built using only those indexes that are average and maximum values of the descriptors of apneas, hypopneas and desaturations. The resulting decision tree has eleven nodes and six leaves and it uses RiseValDesatH, FallValDesatH, FallDurDesatHM and TEndMin.

Table 3 also shows the performance of decision trees built only with the indexes DurFluxM, DurDesatM, DurFlux, DurFluxH, DurDesat, DurDesatHM, DurDesatH, MinValDesat, MinValDesatH, MeanValSpO2 and BasalSpO2. These indexes, which are marked with an asterisk in Table 3, were not selected during the feature selection process. But they are often found in the reports generated by commercial polysomnographs. Therefore we considered it interesting to include them in the table. We also have analyzed the performance of decision trees built with all the other indexes studied in this paper, but we have not included these results in the paper due to their very low prediction capability, which renders them useless for the clinical routine.

## DISCUSSION

According to the results of our study, AHDI is a better tool for discriminating between SAHS and control patients than the AHI. This index outperformed AHI in all the analysis techniques (the different feature evaluation and feature selection techniques, and the classifiers) used in this paper. It also presented the best LR- among the indexes selected in the feature selection process (excluding the decision tree “SAHS if AHI  $> 5$ ”). The criterion “SAHS if AHI  $> 5$ ” (the one recommended by the American Academy of Sleep Medicine<sup>3</sup>) used over our patient database results in perfect sensitivity but low specificity and LR+. Using the criterion “SAHS if AHI  $> 10$ ” we obtain perfect specificity with low sensitivity and high LR-. For any of the three decision trees built with AHI, ADHI provides a better compromise between sensitivity and specificity than AHI; this is reflected in the higher accuracy of the AHDI decision tree when compared with the decision trees built with AHI.

ApHypopPercnt is the index that has the highest LR+ among the indexes selected in the feature selection process, and the highest specificity (excluding the decision tree “SAHS if AHI  $> 10$ ”). However it has a slightly lower sensitivity and accuracy when compared to AHI and AHDI (see Table 3). HypopPercnt presents the best sensitivity after AHDI, but it has a rather poor specificity. DesatPercnt also has good sensitivity, but it has poor specificity.

When building a decision tree using all the indexes, only AHDI was selected. This suggests that once AHDI is known, the other indexes do not add information relevant to the classification problem (as we shall discuss later, this does not mean that they may not be useful to characterize other aspects of the patient’s condition). The decision tree built with all the indexes but AHDI and AHI selected two indexes that convey the same information as AHDI: ApHypopPercnt

and DesatPercent. The classification results of this tree are superior to those of the tree built with AHI. The tree that does not have information about the number of pathological events that have occurred, but only about the characteristics of the events (the “Descriptors” tree) performed poorly (see Table 3).

The AHI only takes into account the number of events that have occurred. Therefore, a patient that experiences 20 apnea events per hour with an average apnea duration of 20 s has exactly the same AHI as a patient that experiences 20 apnea events per hour but whose average event duration is 40 s. However, the latter patient suffers a greater disruption in his/her sleep architecture. Indexes such as AHDI or ApHypopPercent not only take into account the number of events, but also the events’ duration. Therefore, in our previous example both AHDI and ApHypopPercent would take a higher value for the second patient, characterizing more precisely the severity of his/her condition. Considering this temporal information is probably the reason why AHDI outperforms AHI in all metrics, why ApHypopPercent has a better specificity and LR+ than the AHI, and why HypopPercent has a better sensitivity and LR– than the AHI.

DesatPercent and HypopPercent when compared with AHI have the advantage of considering the temporal duration of the pathological events, but have the disadvantage of not considering all types of events that reflect the severity of the patient (the first index only considers desaturations, while second only considers hypopneas). In both cases, the end result is a very good sensitivity at the expense of poor specificity. ApPercent has a worse performance, suggesting that considering only the apneas provides less information than considering only the desaturations or only the hypopneas.

In the medical literature there are a few authors that use indexes similar to ApHypopPercent, ApPercent and DesatPercent.<sup>8,18</sup> However, to the best of our knowledge, a comprehensive evaluation of these indexes has never been carried out. We are not aware of an index similar to AHDI having been used before. All these indexes do not usually appear in the reports of commercial polysomnographs, nor are they among the indexes recommended by the American Academy of Sleep Medicine in their manual.<sup>3</sup>

Patients with SAHS experience repetitive episodes of hypoxia and reoxygenation during transient cessation of breathing that may have systemic effects. These patients also present increased levels of biomarkers linked to endocrine-metabolic and cardiovascular alterations. Moreover, the systemic implication of SAHS may involve sleep fragmentation, tonic elevation of sympathetic neural activity, oxidative stress, inflammation, hypercoagulability and endothelial dysfunction. All of this indicates that SAHS should be

considered a systemic disease rather than a local abnormality. Furthermore, Flemons observed that sinus pauses and bradyarrhythmias were clearly more frequent than normal in patients with SAHS.<sup>11</sup> Some of the indexes studied in this paper may be more suitable for characterizing the long-term risks to the patient’s health, while others may be more suitable for characterizing better the short-term symptoms of the disorder. There is an emerging consensus that SAHS is an oxidative stress disorder. With each apnea, oxygen levels decline and are followed by reoxygenation when breathing resumes. This process augments formation of reactive oxygen species. In turn, increased oxidative stress has been associated with the development of cardiovascular diseases and can be promoted by the chronic intermittent hypoxia characteristic of SAHS.<sup>29</sup> Given the importance of hypoxia in the genesis of cardiovascular problems, it is very important to see if any of the indexes that we have studied can serve as a marker of developing cardiovascular complications. Indexes such as AHDI, ApHypopPercent, HypopPercent or ApPercent may characterize short-term symptoms better than AHI, since they take into account not only the number of respiratory pauses, but also their duration. On the contrary, it is known that nocturnal hypoxaemia can be a major determinant of excessive daytime sleepiness in SAHS patients.<sup>21</sup> Nocturnal hypoxaemia could be quantified by indexes such as DesatPercent, DeltaSpO<sub>2</sub> or AreaSpO<sub>2</sub>. Further research along these lines is required to test these hypotheses.

Besides the AHI, some of the indexes that are more often present in the reports of commercial polysomnographs are the average and maximum duration of apneas, hypopneas and desaturations. None of these indexes made it to the global score of Table 1. Only the maximum durations hypopneas were chosen by one of three feature selection methods (see Table 2). This result was surprising both for the authors and for the team of clinicians which cooperates with us. Our intuition told us that it was not a good idea to characterize the state of the patient using the maximum duration of an event, since that metric would be very sensitive to the occurrence of an abnormally long event. This index has a reasonable sensitivity, but its specificity is very low (see Table 3). All the decision trees built with the average and maximum durations of apneas, hypopneas and desaturations, except the one built with the maximum duration hypopneas, have a very poor classification performance (see Table 3).

The minimum value of SpO<sub>2</sub> during the desaturations is often present in the reports of commercial polysomnographs. However, these reports do not usually distinguish between the desaturations caused by apneas or by hypopneas. The magnitude of the rise



and fall sections of the desaturations caused by hypopneas, which provide information similar to that of the minimum and mean value of SpO<sub>2</sub>, made it to the global score of Table 1. The magnitude of the rise sections of the desaturations caused by hypopneas was chosen by one of the three feature selection methods (see Table 2). The decision trees built with these indexes have an acceptable sensitivity, but very poor specificity (see Table 3). In all the cases, the decision trees built over indexes related to hypopneas performed better than the equivalent decision tree built over the index related to apneas.

DeltaSpO<sub>2</sub> was selected in the feature selection process. Its decision tree has an acceptable sensitivity but low specificity (see Table 3). We have never seen this index in the reports of commercial polysomnographs. However, the two indexes from which it is derived, MeanValSpO<sub>2</sub> and BasalSpO<sub>2</sub>, usually are present. These two indexes were not selected by any feature selection technique and the classifiers built with them are unable to distinguish between the two classes (see Table 3).

## CONCLUSIONS

According to the results of our study, the index AHDI is more useful to discern between healthy and SAHS patients than the AHI. ApHypopPercent, DesatPercent, HypopPercent and DeltaSpO<sub>2</sub> may also provide useful information in diagnosing SAHS. Currently, these indexes are either underutilized or not used at all when assessing SAHS patients' condition. Comprehensive studies on how to take advantage of them, beyond the analysis presented in this paper, have not been carried out. These indexes need to be studied thoroughly since they may provide new tools to assist clinicians in diagnosing SAHS.

Among the indexes that are present more often in the reports of commercial polysomnographs, only the maximum duration of hypopneas and the minimum value of SpO<sub>2</sub> during the saturations (especially those caused by hypopneas) seem to provide a moderate amount of information on the patient's condition. However, the reports contain a large number of indexes that, according to our study, provide little or no information about the severity of the patient's condition. They probably have been included in the reports guided more by marketing reasons ("to have a product with more features"), than because they are of real value to clinicians. To provide useless or low value information increases the cognitive overload of the clinicians, decreasing their efficiency in processing the information that is really useful. Clinicians would benefit from shorter polysomnographic reports, with a

smaller number of indexes, and from having clear guidelines on how to interpret the meaning of the indexes contained in these reports.

## APPENDIX: DATA USED IN THIS PAPER

For the sake of reproducibility of the results, and to allow other researchers to take advantage of the data, we have made publicly available the database used in this paper. Comma separated files containing the 46,505 pathological events, and the 274 feature vectors generated from the previous file, can be found in Otero *et al.*<sup>27</sup> Each of the 274 feature vectors representing patients contains the 42 indexes analyzed in this paper. In addition, each of these feature vectors also contains the patient's age, weight, BMI, and sex. Finally, these feature vectors also have a unique patient identifier, and a diagnosis—class—that can take the values "SAHS" or "Control". The website<sup>27</sup> also contains scripts for the R software environment for statistical computing used in the analysis presented in this paper.

## ACKNOWLEDGMENTS

This work was supported by the Spanish MEC and the European FEDER under the Grant TIN2009-14372-C03-03, by the Xunta de Galicia under the Grant INCITE08SIN002206PR and by the University San Pablo-CEU under the Grant USP BS PPC05/2010.

## REFERENCES

- <sup>1</sup>Almuallim, H., and T. G. Dietterich. Learning with many irrelevant features. In: Proceedings of the Ninth National Conference on Artificial Intelligence, Vol. 2, 1991, pp. 547–552.
- <sup>2</sup>Álvarez-Estévez, D., and V. Moret-Bonillo. Fuzzy reasoning used to detect apneic events in the sleep apnea-hypopnea syndrome. *Expert Syst. Appl.* 36(4):7778–7785, 2009.
- <sup>3</sup>American Academy of Sleep Medicine Task Force. Sleep-related breathing disorders in adults: recommendations for syndrome definition and measurement techniques in clinical research. *Sleep* 22:667–689, 1999.
- <sup>4</sup>Bystricky, W., and A. Safer. Identification of individual sleep apnea events from the ECG using neural networks and a dynamic Markovian state model. In: Computers in Cardiology, 2004, pp. 297–308.
- <sup>5</sup>Cabrero-Canosa, M., M. Castro-Pereiro, M. Graña-Ramos, E. Hernández-Pereira, V. Moret-Bonillo, M. Martín-Egaña, and H. Vereá. An intelligent system for the detection and interpretation of sleep apneas. *Expert Syst. Appl.* 24:335–349, 2003.
- <sup>6</sup>Ciftci, T. U., O. Kokturk, and S. Ozkan. Apnea-hypopnea indexes calculated using different hypopnea definitions and

- their relation to major symptoms. *Sleep Breathing* 8(3): 141–146, 2004.
- <sup>7</sup>Durán, J., S. Esnaola, R. Rubio, and A. Izueta. Obstructive sleep apnea-hypopnea and related clinical features in a population-based sample of subjects aged 30 to 70 yr. *Am. J. Respir. Crit. Care Med.* 163(3):685, 2001.
- <sup>8</sup>Falabella, V., M. Lewis, and C. Campagna. Development of cardiorespiratory patterns associated with terrestrial apneas in free-ranging southern elephant seals. *Physiol. Biochem. Zool.* 72(1):64–70, 1999.
- <sup>9</sup>Fayyad, U., and K. Irani. Multi-interval discretization of continuous-valued attributes for classification learning. In: 13th International Joint Conference on Artificial Intelligence, 1993, pp. 1022–1027.
- <sup>10</sup>Flemons, W. W. Obstructive sleep apnea. *N. Engl. J. Med.* 347(7):498–504, 2002.
- <sup>11</sup>Flemons, W. W., J. E. Remmers, and A. M. Gillis. Sleep apnea and cardiac arrhythmias. is there a relationship? *Am. Rev. Respir. Dis.* 148(3):618, 1993.
- <sup>12</sup>Ghazanshahi, S. D., and M. C. K. Khoo. Optimal ventilatory patterns in periodic breathing. *Ann. Biomed. Eng.* 21(5):517–530, 1993.
- <sup>13</sup>Hall, M. A. Correlation-based feature selection for machine learning. PhD thesis, The University of Waikato, 1999.
- <sup>14</sup>Hearst, M. A., S. T. Dumais, E. Osman, J. Platt, and B. Scholkopf. Support vector machines. *Intell. Syst. Appl. IEEE* 13(4):18–28, 1998.
- <sup>15</sup>Khandoker, A. H., and M. Palaniswami. Modeling respiratory movement signals during central and obstructive sleep apnea events using electrocardiogram. *Ann. Biomed. Eng.* 39(2):801–811, 2011.
- <sup>16</sup>Kira, K., and L. A. Rendell. A practical approach to feature selection. In: Proceedings of the Ninth International Workshop on Machine Learning. Morgan Kaufmann Publishers Inc., 1992, pp. 249–256.
- <sup>17</sup>Kohavi, R., and G. H. John. Wrappers for feature subset selection. *Artif. Intell.* 97(1–2):273–324, 1997.
- <sup>18</sup>Laffont, F., A. Autret, M. Minz, T. Beillevaire, H. P. Cathala, and P. Castaigne. Sleep respiratory arrhythmias in control subjects, narcoleptics and non-cataplectic hypersomniacs. *Electroencephalogr. Clin. Neurophysiol.* 44(6): 697–705, 1978.
- <sup>19</sup>Laxminarayan, P., S. A. Alvarez, C. Ruiz, and M. Moonis. Mining statistically significant associations for exploratory analysis of human sleep data. *IEEE Trans. Inf. Technol. Biomed.* 10(3):440–450, 2006.
- <sup>20</sup>Mantel, N. Chi-square tests with one degree of freedom; extensions of the Mantel-Haenszel procedure. *J. Am. Stat. Assoc.* 58(303):690–700, 1963.
- <sup>21</sup>Mediano, O., A. Barceló, M. de la Peña, D. Gozal, A. Agustí, and F. Barbé. Daytime sleepiness and polysomnographic variables in sleep apnoea patients. *Eur. Respir. J.* 30(1):110, 2007.
- <sup>22</sup>Morsy, A. A., and K. M. Al-Ashmouny. Sleep apnea detection using an adaptive fuzzy logic based screening system. In: 27th IEEE EMB Conference, 2005, pp. 6124–6127.
- <sup>23</sup>Nazeran, H., A. Almas, K. Behbehani, and E. Lucas. A fuzzy inference system for detection of obstructive sleep apnea. In: 23th IEEE EMB Conference, 2001, pp. 1645–1648.
- <sup>24</sup>Otero, A., and P. Félix. Robust Association of Pathological Respiratory Events in SAHS. Lecture Notes in Computer Science, Vol. 5518. Berlin: Springer Verlag, pp. 1020–1027, 2009.
- <sup>25</sup>Otero, A., P. Félix, and M. R. Álvarez. Algorithms for the analysis of polysomnographic recordings with customizable criteria. *Expert Syst. Appl.* 38(8):10133–10146, 2011.
- <sup>26</sup>Otero, A., P. Félix, M. R. Álvarez, and C. Zamarrón. The mean value of the descriptors of the pathological events recorded on the polysomnogram as a support tool in the diagnosis of SAHS. In: 32th Annual International IEEE, EMBS Conference, 2010, pp. 5101–5104.
- <sup>27</sup>Otero, A., P. Félix, and J. M. Rodríguez. Database and r scripts of the paper an evaluation of indexes as support tools in the diagnosis of sleep apnea. <http://biolab.uspcu.com/SAHSIndexEvaluation/>, 2011. Accessed January 2012.
- <sup>28</sup>Otero, A., P. Félix, J. M. Rodríguez, and C. Zamarrón. Evaluation of an alternative definition for the apnea-hypopnea index. In: 32th Annual International IEEE, EMBS Conference, 2010, pp. 4654–4657.
- <sup>29</sup>Park, A. M., H. Nagase, S. V. Kumar, and Y. J. Suzuki. Effects of intermittent hypoxia on the heart. *Antioxid. Redox Signal.* 9(6):723–729, 2007.
- <sup>30</sup>Quinlan, J. R. C4. 5: Programs for Machine Learning. San Mateo: Morgan Kaufmann, 1993.
- <sup>31</sup>Topor, Z. L., M. Pawlicki, and J. E. Remmers. A computational model of the human respiratory control system: responses to hypoxia and hypercapnia. *Ann. Biomed. Eng.* 32(11):1530–1545, 2004.
- <sup>32</sup>van Houdt, P. J., P. P. W. Ossenblok, M. G. van Erp, K. E. Schreuder, R. J. J. Krijn, P. A. J. M. Boon, and P. J. M. Cluitmans. Automatic breath-to-breath analysis of nocturnal polysomnographic recordings. *Med. Biol. Eng. Comput.* 49(9):1–12, 2011.
- <sup>33</sup>Varaday, P., T. Micsik, S. Benedeck, and Z. Benyo. A novel method for the detection of apnea and hypopnea events in respiration signals. *IEEE Trans. Biomed. Eng.* 33:936–942, 2002.
- <sup>34</sup>Young, T., M. Palta, J. Dempsey, J. Skatrud, S. Weber, and S. Badr. The occurrence of sleep-disordered breathing among middle-aged adults. *N. Engl. J. Med.* 328(17):1230–1235, 1993.
- <sup>35</sup>Young, T., P. E. Peppard, and D. J. Gottlieb. Epidemiology of obstructive sleep apnea: a population health perspective. *Am. J. Respir. Crit. Care Med.* 165(9):1217, 2002.
- <sup>36</sup>Young, T., J. Skatrud, and P. E. Peppard. Risk factors for obstructive sleep apnea in adults. *JAMA* 291(16):2013, 2004.