31st Annual International Conference of the IEEE EMBS
Minneapolis, Minnesota, USA, September 2-6, 2009

Screening of patients with Obstructive Sleep Apnea Syndrome using C4.5 algorithm based on non linear analysis of respiratory signals during sleep

Evangelos Kaimakamis, Charalambos Bratsas, *Member, IEEE*, Lazaros Sichletidis, Charalambos Karvounis and Nikolaos Maglaveras, *Senior Member, IEEE*

Abstract—**Aim: To classify patients with possible diagnosis of Obstructive Sleep Apnea Syndrome (OSAS) into groups according to the severity of the disease using a decision tree producing algorithm based on nonlinear analysis of 3 respiratory signals instead of the use of full polysomnography.**

Patients-Methods: Eighty-six consecutive patients referred to the Sleep Unit of a Pulmonology Department underwent full polysomnography and their tests were manually scored. Three nonlinear indices (Largest Lyapunov Exponent-LLE, Detrended Fluctuation Analysis-DFA and Approximate Entropy-APEN) were extracted from two respiratory signals (nasal cannula flow-F and thoracic belt-T). The oxygen saturation signal (SpO_2) was also selected. The above measurements provided data to the C4.5 algorithm using a data mining application.

Results: Two decision trees were produced using linear and nonlinear data from 3 respiratory signals. The discrimination between normal subjects and sufferers from OSAS presented an accuracy of 84.9% and a recall of 90.3% using the variables age, sex, DFA from F and Time with SpO_2<90% (T90). The classification of patients into severity groups had an accuracy of 74.2% and a recall of 81.1% using the variables APEN from F, DFA from F and T90.

Conclusion: It is possible to have reliable predictions of the severity of OSAS using linear and nonlinear indices from only two respiratory signals during sleep instead of performing full polysomnography. The proposed algorithm could be used for screening patients suspected to suffer from OSAS.

I. Introduction

OBSTRUCTIVE Sleep Apnea Syndrome (OSAS) is a common disorder that affects 4% and 2% of middle-aged men and women respectively [1]. The importance of a normal sleep pattern can easily be understood when deteriorations in health status and quality of life [2] of patients suffering from OSAS are taken into consideration. Early screening and detection of the syndrome is mandatory in order to permit the effective therapy with the application of a continuous positive airway pressure (CPAP) appliance [3]. The full polysomnographic study conducted in specially designed sleep laboratories is the gold standard for the proper diagnosis, whereas various techniques such as night oximetry have been utilized to provide an alternative screening test for potential patients, as the number of those subjects is constantly increasing. The effectiveness of these screening techniques has been an issue of controversy [3].

Sleep itself is an active and regulated process which modulates autonomous nervous system functions such as temperature, respiration, blood pressure, and heart rate [4]. Since the regulation of this autonomic activity has been found to be a nonlinear deterministic behavior [5], various researchers have tried to apply measures of nonlinear dynamics to the electroencephalographic (EEG) and electrocardiographic (ECG) signals included in polysomnography [6]. Little has been done to explore these dynamics in pathological respiratory signals in patients with OSAS or apply them to a novel method of screening for the existence of the disease.

Another open issue is the choice of a suitable classification algorithm for the detection of the syndrome of interest. Recently, Polat *et al*, presented comparison of Different Classifier Algorithms on the Automated Detection of Obstructive Sleep Apnea Syndrome [7]. The obtained results have shown that the best classifier system for the diagnosis of obstructive sleep apnea syndrome is *C4.5 decision tree classifier*. Based on that fact, we have tried to integrate the findings of nonlinear analysis of the respiratory signals in polysomnography into a data mining application that produces C4.5 decision tree algorithms. The aim was to classify patients with possible diagnosis of OSAS into groups according to the severity of the disease using a decision tree producing algorithm based on nonlinear analysis of only 3 respiratory signals instead of the use of full polysomnography.

Manuscript received April 23, 2009. This work was supported in part by an educational Grant by Greek State Scholarships Foundation.

Evangelos Kaimakamis is with the Lab of Medical Informatics, Medical School, Aristotle University of Thessaloniki, Greece (corresponding author , phone: +306944-876277; fax: +302310-334710; e-mail: vkaimak@med.auth.gr)

Charalambos Bratsas is with the Lab of Medical Informatics, Medical School, Aristotle University of Thessaloniki, Greece (e-mail: mpampis@ med.auth.gr)

Lazaros Sichletidis is the Head of Pulmonology Department, Aristotle University of Thessaloniki, Greece (e-mail: sichlet@med.auth.gr)

Charalambos Karvounis is with the 1st Cardiology Department, AHEPA General Hospital of Thessaloniki, Greece (e-mail: xaris_karvounis@yahoo.gr)

Nikolaos Maglaveras is with the Lab of Medical Informatics, Medical School, Aristotle University of Thessaloniki, Greece (e-mail: nicmag@ med.auth.gr)

II. METHODS

A. Recording & Pre-processing

Eighty-six consecutive patients referred to the Sleep Unit of the 2nd Pulmonology Department of "G Papanikolaou" Hospital and who accepted to sign the informed consent form were included in the study. The Study had the approval by the ethics committee of the hospital. All the subjects reported symptoms consistent with OSAS and had no significant commorbidities. The subjects underwent overnight attended polysomnography (Somnologica 7000, Flaga; Iceland), including electrocardiography, electroencephalography (C3-A2, C4-A1 leads), electrooculography (EOL, EOR), submental and tibialis electromyography for sleep staging according to standard criteria [8] and respiratory recordings of thoracic and abdominal movements, nasal flow by pressure cannula, snoring, and arterial oxygen saturation with a finger probe using pulse oximetry. Apnea and hypopnea were defined in accordance with standard used criteria [9]. The recordings were manually scored by an experienced medical doctor.

In total 24 subjects were found to be normal (Apnea-Hyponea Index-AHI less than 5/hour) and 62 suffered from OSAS. The severity of the syndrome was classified as follows: 5<AHI≤15=mild, 15<AHI≤30=moderate and AHI>30=severe. Table I summarizes the descriptive statistical data from the total 86 patients.

Three nonlinear indices (Largest Lyapunov Exponent-LLE, Detrended Fluctuation Analysis-DFA and Approximate Entropy-APEN) were extracted from two respiratory signals (nasal cannula flow-F and thoracic belt movement-T). The oxygen saturation signal (SpO_2) from pulse oximetry was also selected. The above signals had a mean duration of 315 minutes and were first exported in European data Format to be further processed with the use of signal processing software (Matlab by Mathworks Inc.) in personal computers. The LLE calculation required the use of a command line application by Rosenstein *et al* as well as a spreadsheet program (Microsoft Excel.)

The basic statistical analysis was performed with the use of SPSS for Windows, Version 15.0 (SPSS Inc, Chicago, Illinois).

TABLE I
DESCRIPTIVE STATISTICS

	N	Mean	Std. Deviation
AGE	86	47,73	13,44
EPWORTH	79	8,27	5,26
BMI	82	32,33	6,56
NEC_CIRC	64	41,56	4,71
WAIST	63	110,71	17,25
HIP	64	114,41	14,30
TST	86	314,99	58,67
T90	86	23,08	29,43
AHI	86	35,37	33,89
AI	86	24,20	30,88
HI	86	11,13	11,23
LLEf	86	0,64	0,61
LLEt	79	1,12	0,81
DFA slow_f	86	0,28	0,29
DFA fast_f	86	0,32	0,34
DFA slow_t	79	0,18	0,23
DFA fast_t	79	0,55	0,18
APEN low_f	86	-8,77	9,48
APEN high_f	86	8,73	8,89
APEN low_t	79	-29,37	32,76
APEN high_t	79	33,81	30,97

BMI=Body Mass Index, NEC_CIRC=Neck circumference 9(in cm), TST=Total Sleep Time (in min), T90=Time with SaO_2<90%, AHI=Apnea-Hypopnea Index, AI=Apnea Index, HI=Hypopnea Index, LLE=Largest Lyapunov Exponent, f=flow signal, t=thoracic belt signal, DFA=Detrended Fluctuation Analysis α factor (slow-fast), APEN=Approximate Entropy

B. Feature extraction

As concerns the methods of analysis that were selected, the following information is explanatory. The method of detrended fluctuation analysis [10] has proven useful in revealing the extent of long-range correlations in time series. Briefly, the time series to be analyzed (with N samples) is first integrated. Next, the integrated time series is divided into boxes of equal length, n. In each box of length n, a least squares line is fit to the data (representing the trend in that box). The y coordinate of the straight line segments is denoted by yn(k).

Next, we detrend the integrated time series, y(k), by subtracting the local trend, yn(k), in each box. The root-mean-square fluctuation of this integrated and detrended time series is calculated by

$$F(n) = \sqrt{\frac{1}{N}\sum_{k=1}^{N}[y(k) - y_n(k)]^2}.$$

This computation is repeated over all time scales (box sizes) to characterize the relationship between F(n), the average

fluctuation, and the box size, n. Typically, F(n) will increase with box size. A linear relationship on a log-log plot indicates the presence of power law (fractal) scaling. Under such conditions, the fluctuations can be characterized by a scaling exponent, the slope of the line relating log F(n) to log n. The boxes selected for the signal analyses included time periods of 8, 30, and 100 seconds, which represent the mean duration of 2, 6 and 24 breaths respectively. The time series used for the calculation of DFA from F were 20 minutes long and the same duration for the T signals was 240 minutes. The difference in the length of the two types of signals reflects the limitations in the computational capabilities of the personal computers used, due to the large overall size of the flow signals (as a result from the 200Hz sampling in this case). Two DFA measurements were obtained from each signal, the DFA fast value which represented the power law slope on the medium to fast time scales, as well as the DFA slow value, i.e. the slope on the slow to medium time scales.

Entropy, as it relates to dynamical systems, is the rate of information production. Methods for estimation of the entropy of a system represented by a time series are not, however, well suited to analysis of the short and noisy data sets encountered in cardiovascular and other biological studies. Pincus introduced approximate entropy (ApEn) [11], a set of measures of system complexity closely related to entropy, which is easily applied to clinical cardiovascular and other time series. The method examines time series for similar epochs: more frequent and more similar epochs lead to lower values of ApEn. Informally, given N points, the family of statistics ApEn(m, r, N) is approximately equal to the negative average natural logarithm of the conditional probability that two sequences that are similar for m points remain similar, that is, within a tolerance r, at the next point. Thus a low value of ApEn reflects a high degree of regularity. Importantly, the ApEn algorithm counts each sequence as matching itself, a practice carried over from the work of Eckmann and Ruelle to avoid the occurrence of ln (0) in the calculations [12]. The values selected in our study were: m=2, r=0.2, N=the total sleep recording. Alternative values for these parameters showed no significant alteration in the derived APEN figures.

The Largest Lyapunov exponent estimation was based on the complexity theory: Consider two points in a space: X_0 & $X_0 + Dx_0$, each of which will generate an orbit in that space using some equation or system of equations. These orbits can be thought of as parametric functions of a variable like time. If we use one of the orbits as a reference orbit, then the separation between the two orbits will also be a function of time. Because sensitive dependence can arise only in some portions of a system (like the logistic equation), this separation is also a function of the location of the initial value and has the form $Dx(X_0, t)$. In a system with attracting fixed points or attracting periodic points, $Dx(X_0, t)$ diminishes asymptotically with time. If a system is unstable, like pins balanced on their points, then the orbits diverge exponentially for a while, but eventually settle down. For chaotic points, the function $Dx(X_0, t)$ will behave erratically. It is thus useful to study the mean exponential rate of divergence of two initially close orbits using the formula:

$$\lambda = \lim_{\substack{t\to\infty \\ |\Delta x_0|\to 0}} \frac{1}{t} \ln \frac{|\Delta x(X_0,t)|}{|\Delta x_0|}$$

This number, called the Lyapunov exponent "l" [lambda], is useful for distinguishing among the various types of orbits. It works for discrete as well as continuous systems. Various techniques have been developed to calculate this measure. One of the most robust and practical is the method proposed by Rosenstein *et al* [13]. The method is suitable for calculating the LLE from small data sets with reliability. For the flow signals, the method was applied for periods of 6 minutes, whereas 170-minute-periods were analyzed for the thoracic belt signals.

C. *Feature Selection phase*

Feature selection is one of the most important steps in pattern recognition or pattern classification and data mining. It is difficult to measure classification information in all features [14]. Data preprocessing is an indispensable step in effective data analysis. It prepares data for data mining and machine learning, which aim to turn data into business intelligence or knowledge. Feature selection is a data preprocessing technique commonly used on high dimensional data. Feature selection studies how to select a subset or list of attributes or variables that are used to construct models describing data.

Feature selection is normally done by searching the space of attribute subsets, evaluating each one. This is achieved by combining attribute subset evaluator with a search method.

For the feature selection phase, two objects must be set up: a feature evaluator and a search method. The evaluator determines what method is used to assign a worth to each subset of features. The search method determines what style of search is performed.

The feature selection can be done in two ways: 1) using full training set (the worth of the feature subset is determined using the full set of training data), or 2) by cross-validation (the worth of the feature subset is determined by a process of cross-validation). In addition, the classifying time grows dramatically with the number of features, rendering the algorithm impractical for problems with a large number of features.

In this study, we have chosen the *classifier subset evaluator* as a feature evaluator and *BestFirst* as a search method. The *classifier subset evaluator* evaluates attribute subsets on training data or a separate hold out testing set. Furthermore, we have used the J48 classifier as a classifier to estimate the 'merit' of a set of attributes. The *BestFirst* searches the space of feature subsets by greedy hill-climbing augmented with a backtracking facility.

D. Classification Techniques

C4.5 Decision tree learning is one of the most widely used and practical methods for inductive inference. It is a method for approximating discrete-valued functions that is robust to noisy data and capable of learning disjunctive expressions [15, 16]. C4.5 decision tree learning is a method for approximating discrete-valued functions, in which the learned function is represented by a decision tree. Learned trees can also be represented as sets of if–then rules to improve human readability. These learning methods are among the most popular of inductive inference algorithms and have been successfully applied to a broad range of tasks including learning to diagnose medical cases. C4.5 Decision tree learning is a heuristic, one-step look ahead (hill climbing), non-backtracking search through the space of all possible decision trees [15–17].

The aim of C4.5 decision tree learning is to recursively partition data into sub-groups. Working of C4.5 decision tree learning is as follows:

Select an attribute and formulate a logical test on attribute

- Branch on each outcome of test, move subset of examples (training data) satisfying that outcome to the corresponding child node
- Run recursively on each child node
- Termination rule specifies when to declare a leaf node

Overfitting was avoided by evaluating the classification algorithms using 10-fold cross-validation.

The performance of each classifier was assessed with a stratified 10-fold cross-validation method. This approach has the advantage that all the data is, at some point, used for model evaluation, as opposed to simply splitting the data into testing and training sets. Instead, the data was divided into 10 equal sized fragments, each of which was in turn used as an independent test set, while the other fractions were used for training the classifier. Classification error was then estimated as the average performance over the 10 test sets. The classification and feature selection tasks were performed with the help of a freely available software package Weka [18, 19], Version 3.5.8 (Weka Machine Learning Project, The University of Waikato, New Zealand) by calling it from R.

III. Results

Two decision trees were produced using linear and nonlinear data from the three respiratory signals and supplying them to the data mining application mentioned above. The fist decision tree presents an algorithm to discriminate between normal subjects and patients with OSAS, whereas the second one is used to categorize patients into groups of disease severity.

The discrimination between normal subjects and sufferers from OSAS presented an accuracy of 84.9% and a recall of 90.3% (Method statistics can be seen in Table II) using the variables age, sex, DFA from nasal cannula flow (F) and Time with SpO_2<90% (T90). It is interesting that only two respiratory signals are exploited in this case providing sufficient data to screen for sleep apneas.

TABLE II
C4.5 CLASSIFICATION TREE STATISTICS

Correctly Classified Instances	73	84.89 %
Incorrectly Classified Instances	13	15.11 %
Kappa statistic	0.6195	
Mean absolute error	0.2082	
Root mean squared error	0.3863	
Relative absolute error	51.37 %	
Root relative squared error	86.03%	
Total Number of Instances	86	

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
0.708	0.097	0.739	0.708	0.723	0.782	Normal
0.903	0.292	0.889	0.903	0.896	0.782	Abnormal

The second produced algorithmic sequence is suitable for the classification of OSAS patients into groups according to the severity of the syndrome. The following figure 1 shows the produced decision tree for the classification.

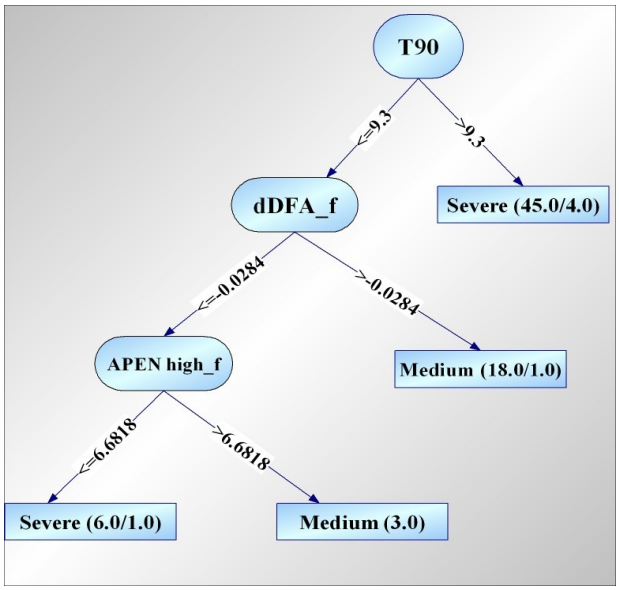

Fig.1. Desicion tree produced by C4.5 algorithm for the classification of OSAS patients into severity groups.

The classification of patients into severity groups had an accuracy of 74.2% and a recall of 81.1% using the variables APEN from F, DFA from F and T90 (Table III). Again we have found that the use of only two signals (nasal flow and pulse oximetry) are adequate for the estimation of the severity level. Moreover the algorithm was more precise in cases with severe OSAS, which will definitely need the use of CPAP devices for therapeutic purposes.

TABLE III
C4.5 STATISTICS FOR SEVERITY OF OSAS

Correctly Classified Instances	46	74.19%
Incorrectly Classified Instances	16	25.81 %
Kappa statistic	0.4567	
Mean absolute error	0.3021	
Root mean squared error	0.4954	
Relative absolute error	62.56 %	
Root relative squared error	100.77 %	
Total Number of Instances	62	

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
0.811	0.36	0.769	0.811	0.789	0.645	Severe
0.64	0.189	0.696	0.64	0.667	0.645	Medium

IV. CONCLUSION

It is possible to have reliable predictions of the severity of OSAS using linear (like the Oxygen saturation) and nonlinear indices (like DFA and APEN) from only two respiratory signals (Flow from nasal cannula and pulse oximetry) during sleep instead of performing full polysomnography. The selected signals are reliably indicative of disturbed respiration during sleep and are universally considered easy and practical to use [20]. The utilization of those biosignals alone for accurate prediction of sleep apneas offers an obvious advantage to clinicians and sleep researchers, as it can alternatively be used instead of the expensive and time consuming full polysomnography. Furthermore, contemporary portable devices that can measure these parameters are currently under development and are expected to bring a new era of remote signal acquisition in home care applications [21]. The integration of the proposed algorithms in such devices will certainly boost screening for OSAS in remote areas and without the need for large specialized sleep units, thus enhancing telemedicine capabilities.

The proposed algorithms could therefore be used for screening patients suspected to suffer from OSAS. The obtain results show promising levels of accuracy in this field and the equipment needed to perform this screening is generally cheap and easy to operate.

Future similar studies with the use of more powerful computers are required in order to explore the effect of analyzing longer time series on the precision and recall features of the methodology. In addition, further analysis with a larger number of patients could show the effectiveness and reproducibility of the proposed method and explore the trends in the precision and recall characteristics in groups of patients with other commorbidities, like congestive heart failure or overlap syndrome.

REFERENCES

[1] W. T. McNickolas, M. R. Bonsigore, "Sleep apnea as an independent risk factor for cardiovascular disease: current evidence, basic mechanisms and research priorities," Eur Respir J 2007; 29:156-178.

[2] V. Tsara, E. Kaimakamis, E. Serasli, Z. Katsarou, P. Christaki, "Health related quality of life in Greek patients with sleep apnea-hypopnea syndrome treated with continuous positive airway pressure," Sleep Med. 2009 Feb; 10(2):217-225.

[3] A. Minaritzoglou, E. Vagiakis, "Polysomnography: Recent data on procedure and analysis," Pneumon 2008;21(4):348-368.

[4] T. Penzel, N. Wessel, M. Riedl, J. W. Kantelhardt, S. Rostig, M. Glos, A. Suhrbier, H. Malberg, I. Fietze, "Cardiovascular and respiratory dynamics during normal and pathological sleep," Chaos 17, 015116(2007).

[5] C-K Peng, S. Havlin, H.E. Stanley, A.L. Goldberger, "Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series.," Chaos 1995;5:82-87.

[6] L.A. Aguirre, V.C. Barros, A.V. Souza, "Nonlinear multivariable modeling and analysis of sleep apnea time series," Comput Biol Med. 1999 May;29(3):207-28.

[7] K. Polat, Ş. Yosunkaya, S. Güneş, "Comparison of Different Classifier Algorithms on the Automated Detection of Obstructive Sleep Apnea Syndrome," Journal of Medical Systems, 33(3),pp. 243-250, 2008.

[8] A. Rechtschaffen, A. Kales, "A manual of standardized terminology techniques and scoring system for sleep stages of human subjects," Washington, DC: National Institute of Health; 1968

[9] American Academy of Sleep Medicine (AASM), "Sleep related breathing disorders in adults: recommendations for syndrome definition and measurement techniques in clinical research," Sleep1999;22:667–89.

[10] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. Ch. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C-K Peng, H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals," Circulation 101(23):e215-e220 [Circulation Electronic Pages; http://circ.ahajournals.org/cgi/content/full/101/23/e215]; 2000.

[11] S. M. Pincus, "Approximate entropy (ApEn) as a complexity measure," Chaos 5: 110-117, 1995.

[12] J. S. Richman and J. R. Moorman. "Physiological time series analysis using approximate entropy and sample entropy," Am J Physiol 2000; 278(6):H2039-H2049.

[13] M.T. Rosenstein, J.J. Collins, and C.J. De Luca, "A practical method for calculating largest Lyapunov exponents from small data sets," Physica D 65:117-134, 1993.

[14] B. Krishnapuram, A. J. Hartemink, L. Carin and M. A.T. Figueiredo, "A Bayesian Approach to Joint Feature Selection and Classifier Design", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 26, No. 9,2004, pp1105 – 1111.

[15] M. T. Mitchell, "Machine learning," McGraw-Hill, Singapore, 1997.

[16] J. R. Quinlan, "Induction of decision trees," Mach. Learn 1:81–106, 1986.

[17] K. Polat, and S. Günes, "The effect to diagnostic accuracy of decision tree classifier of fuzzy and k-NN based weighted pre-processing methods to diagnosis of erythemato-squamous diseases," Digit. Signal Process 16(6):922–930, 2006.

[18] I.H. Witten and E. Frank, "Data mining: practical machine learning tools and techniques", 2nd ed., San Francisco: Morgan Kaufmann; 2005.

[19] Delen D, Walker G, Kadam A. Predicting breast cancer survivability: a comparison of three data mining methods. Artif Intell Med 2005; 34:113—27.

[20] N. J. Douglas, "Clinician's guide to sleep medicine," London, UK: Arnold, 2002, pp. 82–83.

[21] Information Society Technologies Program "Sensation" Deliverable D3.1.3: "Updated user requirements, applications and medical protocols; list of policies and procedures to be followed," 2007.