AdaBoost Classification to Detect Sleep Apnea from Airflow Recordings

G.C. Gutiérrez-Tobal[1], D. Álvarez[1], J. Gómez-Pilar[1], F. del Campo[1,2], and R. Hornero[1]

[1] Grupo de Ingeniería Biomédica, ETSI Telecomunicación, Universidad de Valladolid, Valladolid, España
[2] Hospital Universitario Rio Hortega, Valladolid, España

Abstract—**In this paper, we focus on the automatic detection of sleep apnea-hypopnea syndrome (SAHS) from single-channel airflow (AF) recordings. Spectral data from a very low frequency band of AF is used to feed classifiers based on linear discriminant analysis (LDA). These are iteratively obtained through the *AdaBoost.M1* (ABM1) algorithm, which combines their performance in order to reach a higher diagnostic ability. We built an ABM1-LDA model, using a training set, which showed generalization ability as well as high diagnostic statistics in an independent test set (94.1% sensitivity, 85.7% specificity, and 92.7% accuracy). These results outperform those from recent studies focused on scoring apneas and hipopneas. Hence, the utility of our approach to assist in SAHS diagnosis is showed.**

Keywords—**sleep apnea hypopnea syndrome, airflow, spectral analysis, linear discriminant analysis, boosting.**

I. INTRODUCTION

The sleep apnea-hypopnea syndrome (SAHS) is a prevalent illness that affects both health and life quality of diseased [1]. Patients suffering from SAHS experiment recurrent episodes of complete cessation (apneas) and significant reduction (hypopneas) of airflow (AF) during sleep [2]. Apneas and hypopneas lead to oxygen desaturations and arousals [3], which avoid resting while sleeping. Inadequate rest derives in poor life quality due to daytime symptoms such as hypersomnolence, cognitive impairment, and depression [1]. Some of them have been related to occupational accidents and motor vehicle collisions [4], [5]. Moreover, SAHS is usually associated with major cardiovascular diseases such as stroke, myocardial infarction, cardiac failure, and hypertension [3]. Recently, it has been also related to an increase in cancer incidence [6].

The "gold standard" for SAHS diagnosis is overnight polysomnography (PSG) [7]. Despite its effectiveness, the PSG is a complex test since it requires monitoring and recording multiple physiological signals from subjects during sleep, such as electroencephalogram (EEG), electrocardiogram (ECG), oxygen saturation of hemoglobin (SpO_2) or airflow (AF) [7]. It is also costly due to the expensive equipments and specialized workforce required for the acquisition of the signals and the patients' care, respectively [2]. Furthermore, physicians must perform an offline inspection of the recordings in order to derive the apnea-hypopnea index (AHI), the parameter used to establish SAHS severity. Hence, PSG is also time-consuming [2]. These drawbacks, in turn, lead to increased time to reach diagnosis and treatment due to large waiting lists [2].

The search for diagnostic alternatives has mainly relied on studying a reduced set of signals from PSG [2]. The automatic analysis of a single signal has been proposed in order to minimize complexity, cost, and time of the test [2]. This approach facilitates the implementation of diagnostic portable devices. Since AF is directly modified by apneas and hypopneas [8], its investigation is a natural way of dealing with the problem of SAHS detection. A number of studies aimed at diagnosing SAHS in time domain from single-channel AF by automatic scoring of apneas and hypopneas [9]-[11]. Alternatively, due to the recurrence of apneic events, we propose a global analysis of AF recordings in the frequency domain.

Previous studies showed the usefulness of data from the very low frequency band of AF spectrum to help in SAHS diagnosis [12], [13]. Thus, our first step is to characterize this band by the extraction of several spectral features. Then, we propose the use of *AdaBoost.M1* (ABM1) along with linear discriminant analysis (LDA) in order to classify the spectral data. ABM1 is a boosting algorithm commonly used to combine the diagnostic ability of several weak classifiers to reach generalized models [14]. LDA, which acts as the weak classifier, has been already used to detect SAHS from oximetry recordings [15].

In this study, the ability of the proposed methodology to help in SAHS diagnosis is assessed. Our hypothesis is that the information contained in the very low frequency band of single-channel AF can be used along with the generalization ability of ABM1 to accurately detect SAHS.

II. POPULATION AND SIGNAL UNDER STUDY

This study involved overnight AF recordings from 104 subjects: 86 diseased (SAHS-positive) and 18 non-diseased (SAHS-negative). The recordings were acquired during the PSG, which was performed using a polygraph (E-Series, Compumedics) in the sleep unit of the Hospital Universitario Río Hortega (Valladolid, Spain). The sensor used to obtain AF was a nasal prong pressure (NPP) and the sample rate was 128 Hz. All the subjects were suspected of suffering from

L.M. Roa Romero (ed.), *XIII Mediterranean Conference on Medical and Biological Engineering and Computing 2013*, IFMBE Proceedings 41,
DOI: 10.1007/978-3-319-00846-2_451,

SAHS before undergoing PSG due to common symptoms such as daytime sleepiness, loud snoring, nocturnal choking and awakenings, and referring apneic events. Physicians scored apneas and hypopneas according to the American Academy of Sleep Medicine (AASM) rules [8]. They established AHI = 10 events per hour (e/h) as the threshold for a positive diagnosis. Subjects were randomly divided into a training set (60%) and a test set (40%). All of them gave their informed consent to participate in the study and the Review Board on Human Studies accepted the protocol. Table 1 shows demographic and clinical data of the subjects, such as body mass index (BMI) or age (mean ± standard deviation), from the entire set and each group.

Table 1 Demographic and clinical data

	All	SAHS-negative	SAHS-positive
All subjects			
# Subjects	104	18	86
Age (years)	52.1 ± 15.0	42.3 ± 12.8	54.1 ± 14.7
Men (%)	73 (70.2)	12 (66.7)	61 (70.9)
BMI (kg/m^2)	31.2 ± 6.2	29.3 ± 6.1	31.6 ± 6.1
AHI (e/h)	-	6.2 ± 2.3	44.6 ± 26.9
Training set			
# Subjects	63	11	52
Age (years)	51.5 ± 15.6	42.2 ± 13.1	53.6 ± 15.4
Men (%)	44 (69.8)	8 (72.7)	36 (69.2)
BMI (kg/m^2)	30.3 ± 5.9	28.8 ± 6.4	30.7 ± 5.8
AHI (e/h)	-	6.7 ± 1.3	43.6 ± 27.5
Test set			
# Subjects	41	7	34
Age (years)	52.9 ± 14.2	43.6 ± 14.1	54.9 ± 13.7
Men (%)	29 (70.7)	4 (57.1)	25 (73.5)
BMI (kg/m^2)	32.5 ± 6.4	30.5 ± 6.4	32.8 ± 6.4
AHI (e/h)	-	5.2 ± 3.1	46.0 ± 26.3

III. METHODS

Our methodology uses data from the very low frequency components of AF power spectral density (PSD) to feed several LDA classifiers. These were obtained iteratively, through ABM1, in order to combine the diagnostic ability of all classifiers and improve the overall performance.

A. Power Spectral Density and Feature Extraction

A PSD for each AF recording was estimated by the Welch periodogram, which is suitable for non-stationary signals [16]. We used 50% overlap, a Hamming window with 2^{15} samples, and a discrete Fourier transform (DFT) of 2^{16} points. PSDs were normalized (PSDn) by dividing the amplitude at each frequency by the total spectral power of the signal. Figure 1 shows the averaged PSDns for the SAHS-positive and SAHS-negative groups in the training set. As reported in previous studies [12], [13], there exists a PSD increase in the very low frequency components of the SAHS-positive subjects. Since apneic events last 10 seconds or more [8], their corresponding frequency components are located in the range 0.01-0.10 Hz. Hence, we characterized it by extracting five common features:

- First-to-fourth statistical moments in frequency domain (Mf_1-Mf_4), corresponding to mean, standard deviation, skewness, and kurtosis in the band 0.01-0.10 Hz.
- Peak amplitude (PA), which corresponds to the local maximum in the band 0.01-0.10 Hz.

Mf_1-Mf_4 and PA values from each subject were stored into an associated vector, $\mathbf{x}_i \in \mathbf{x}$, with i = 1, 2,…, N, where N is the total number of subjects, and $\mathbf{x}$ is the whole dataset.

B. Linear Discriminant Analysis

LDA is a supervised classifier which assigns data, $\mathbf{x}$, into one out of k classes, C_k. It relies on the assumption that the conditional class density function of each class, $p(\mathbf{x}|C_k)$, follows a multivariate normal distribution (normality), with identical covariance matrices, $\boldsymbol{\Sigma}$, for all the classes (homocedasticity) [17]. A discriminant score $y_k(\mathbf{x})$ is computed for each class following [15]:

$$y_k(\mathbf{x}) = \boldsymbol{\mu}_k^T \boldsymbol{\Sigma}^{-1} \mathbf{x} - \frac{1}{2} \boldsymbol{\mu}_k^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_k + \ln P(C_k), \quad (1)$$

where $\boldsymbol{\mu}_k$ is the mean vector for class C_k and $P(C_k)$ its corresponding prior probability, i.e., the initial proportion of vectors $\mathbf{x}_i$ belonging to class C_k. Since we only consider two classes (SAHS-positive and SAHS negative), the classification task is performed by the decision rule, "assign a new vector $\mathbf{x}_i$ to the class C_k if $y_k(\mathbf{x}_i) = \max_{k=1,2} y_k(\mathbf{x}_i)$ ". In our study, $P(C_k)$, $\boldsymbol{\mu}_k$, and $\boldsymbol{\Sigma}$ were computed from the training set, since they form the LDA model.

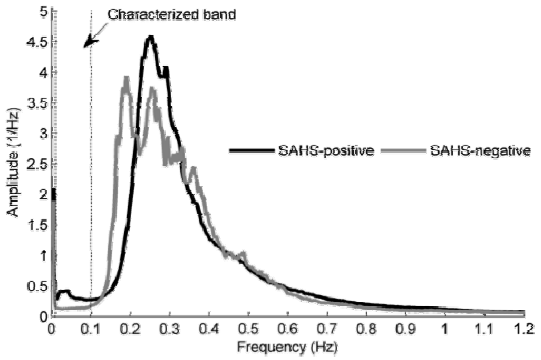

Fig. 1 Averaged PSDn for SAHS-positive and SAHS-negative subjects in the training set

C. AdaBoost.M1

Boosting procedures are iterative algorithms designed to combine models that complement one another [14]. These techniques combine models of the same type using a weighted vote of the prediction from each one [14], [17]. *AdaBoost.M1* (adaptive boosting, ABM1) is a widely used method to perform boosting, originally developed by Freund and Schapire [18]. ABM1 can be used along with any classifier [14]. However, applying ABM1 to too-complex ones could lead to a poor performance when classifying new data [14]. Hence, simple procedures known as weak classifiers, such as LDA, are preferable [14].

ABM1 starts assigning the same weight, w_i, to each instance or vector $\mathbf{x}_i$ in the training set. Typically, $1/N_{tr}$ is used [17], being N_{tr} the number of training instances. The iterative process begins by assessing the performance of the first classifier when assigning weighted instances into the right class. Thus, an error, ε, is computed by dividing the sum of the weights of the misclassified instances by the total weights of all instances [14]. This error is used to determine a weight, α_m, for the current m-classifier following:

$$\alpha_m = \ln\frac{1-\varepsilon_m}{\varepsilon_m} \qquad (2)$$

Then, the weights of the misclassified instances are updated by the expression:

$$w_i^{m+1} = w_i^m \cdot \frac{1-\varepsilon_m}{\varepsilon_m} \qquad (3)$$

Next, all the weights w_i^{m+1} are normalized in order to sum the same as the previous ones, w_i^m, and an additional classifier is assessed using the updated weighted instances [14]. The iterative process automatically ends when $\varepsilon_m = 0$ or $\varepsilon_m \geq 0.5$. The updating of w_i provides higher values to those instances misclassified during the previous iteration [14]. Hence, the new classifier provides these instances with more relevance, being more likely to classify them rightly [14]. The final classification task is performed by returning the class C_k with the highest sum of the votes along all classifiers, weighted by the corresponding α_m value. Thus, those classifiers with smaller ε_m contribute more to the final decision. It has been proved that using the given expressions for ε, w_i, and α_m is equivalent to a sequential minimization of the exponential error function [17].

D. Statistical Analysis

The diagnostic ability of each single spectral features, a conventional LDA classifier, and ABM1-LDA were assessed in terms of sensitivity (Se, proportion of diseased subjects rightly classified), specificity (Sp, proportion of non-diseased subjects rightly classified), and accuracy (Acc, proportion of all subjects rightly classified). To find an optimum threshold, u_o, for the assessment of each single feature, a receiver operating-characteristics (ROC) analysis was done. For each feature, u_o was selected in the training set according to the minimum Euclidean distance between the pair (Se, 1-Sp) and the point (1, 0).

IV. Results

ABM1 iteratively formed five LDA models until satisfying the stopping criterion ($\varepsilon_m = 0$ or $\varepsilon_m \geq 0.5$). These were obtained from the instances in the training set. Table 2 shows the performance of each one, its corresponding ε and α values, and the performance of ABM1. The LDA with the lowest ε value was reached at iteration # 1. Hence, its contribution to the final voting, α, was the highest. None of the single classifiers outperformed ABM1-LDA in terms of accuracy (84.1%). The five LDA models were subsequently applied to the test instances. Their predictions were considered according to each α, in order to perform the final classification. Table 3 summarizes the diagnostic performance of ABM1-LDA, every single feature, and a conventional LDA in the test set. It also shows the optimum thresholds u_o to classify the subjects by the use of single features. As expected, ABM1-LDA generalizes better since it widely outperformed each single feature and LDA. It improved the performance of ABM1-LDA in the training set as well. ABM1-LDA not only reached the highest accuracy (92.7%), but also achieved a balanced sensitivity/specificity pair (94.1% / 85.7%, respectively).

V. Discussion and Conclusions

We obtained an ABM1 model to detect SAHS. It was composed of five LDA classifiers and was developed using features from a very low spectral band in AF (0.01-0.10 Hz). Our model showed high generalization ability, reaching high diagnostic performance when applied to independent test data (94.1% Se, 85.7% Sp, and 92.7% Acc).

Recent studies addressed the automatic diagnosis of SAHS through single-channel AF [9]-[11]. The common aim was to detect and score respiratory events in AF time series. These studies involved from 59 to 131 subjects and reported Se, Sp, and Acc values ranging 80.4-91.5 %, 82.3-87.5 %, and 81.2-89.3 %, respectively [9]-[11]. When using single-channel AF, the event-by-event approach scores AF reductions which are not truly hypopneas, since they should be accompanied by a 3% or more decrease in SpO_2 [8]. Alternatively, the very low spectral band that we used was previously related to desaturations [12]. This could be a reason for our higher results. However, further investigation is required to address this issue.

Table 2 Performance of each LDA classifier and ABM1-LDA (Training set)

		ε	α	Se(%)	Sp(%)	Acc(%)
LDA iteration		0.175	1.553	98.1	1.0	82.5
	# 2	0.355	0.598	63.5	72.7	65.1
	# 3	0.277	0.961	75.0	81.8	76.2
	# 4	0.407	0.378	76.9	36.4	69.8
	# 5	0.372	0.523	69.2	63.6	68.2
ABM1-LDA		-	-	90.4	54.5	84.1

Table 3 Diagnostic ability of single features, LDA, and ABM1-LDA (Test set)

	u_0	Se(%)	Sp(%)	Acc(%)
PA	0.269	76.5	85.7	82.9
Mf_1	0.184	70.6	100.0	75.6
Mf_2	0.040	70.6	85.7	73.2
Mf_3	0.549	82.3	71.4	80.5
Mf_4	2.816	76.5	71.4	75.6
LDA	-	100.0	14.3	85.3
ABM1-LDA	-	94.1	85.7	92.7

In spite of the effectiveness showed by our methodology, some limitations need to be addressed. First, more subjects are required to increase the statistical power of our results. Accordingly, the number of SAHS-negative subjects should be higher. However, our sample reflects a realistic proportion of diseased and non-diseased subjects who undergo PSG. Finally, several weak classifiers could be used along with ABM1 and newer versions of *AdaBoost*. The assessment of an optimum combination of classifiers and boosting algorithms to help in SAHS diagnosis is a future goal.

Summarizing, the usefulness of AF spectral data from very low frequencies was showed. We outperformed recent studies focused on a common event-by-event scoring approach. We also obtained an ABM1-LDA model which achieved generalization ability as well as high accuracy, showing its utility to help in SAHS detection.

ACKNOWLEDGMENT

This research was supported in part by the Proyecto Cero 2011 on Ageing from Fundación General CSIC, Obra Social La Caixa and CSIC, the "Consejería de Educación (Junta de Castilla y León)" under project VA111A11-2, and project TEC2011-22987 from Ministerio de Economía y Competitividad and FEDER. G. C. Gutiérrez-Tobal was in receipt of a PIRTU grant from the Consejería de Educación de la Junta de Castilla y León and the European Social Fund (ESF).

REFERENCES

1. Young T, Peppard PE, Gottlieb DJ (2002) Epidemiology of Obstructive Sleep Apnea: A Population Health Perspective. Am J Respir Crit Care Med 165:1217-39
2. Flemmons WW, Littner MR, Rowley et al. (2003) Home diagnosis of sleep apnea: a systematic review of the literature. Chest 124:1543-79
3. López-Jiménez F, Kuniyoshi FHS, Gami A, et al. (2008) Obstructive sleep apnea: implications for cardiac and vascular disease. Chest 133:793-804
4. Lindberg E, Carter N, Gislason T, et al. (2001) Role of snoring and daytime sleepiness in occupational accidents. Am J Respir Crit Care Med 164:2031-35
5. Sassani A, Findley LJ, Kryger M, et al. (2004) Reducing motor-vehicle collisions, cost, and fatalities by treating obstructive sleep apnea syndrome. Sleep 27:453-58
6. Campos-Rodríguez F, Martínez-García MA, Martínez M, et al. (2013) Association between obstructive sleep apnea and cancer incidence in a large multicenter Spanish cohort. Am J Respir Crit Care Med 187:99-105
7. Patil SP, Schneider H, Schwartz AR, et al. (2007) Adult obstructive sleep apnea: pathophysiology and diagnosis. Chest 132:325-337
8. Iber C, Ancoli-Israel S, Chesson A, et al. (2007) The AASM manual for the scoring of sleep and associated events: rules, terminology and technical specifications (Westchester, IL, USA: American Academy of Sleep Medicine)
9. Erman MK, Stewart D, Einhorn D et al. (2007) Validation of the ApneaLinkTM for the screening of sleep apnea: a novel and simple single-channel recording device. J Clin Sleep Med 3:387-392
10. Ragette R, Wang Y, Weinreich G et al. (2010) Diagnostic performance of single airflow channel recording (ApneaLink) in home diagnosis of sleep apnea. Sleep Breath 14:109-14
11. Nigro CA, Dibur E, Aimaretti S et al. (2011) Comparison of the automatic analysis versus the manual scoring from ApneaLink™ device for the diagnosis of obstructive sleep apnoea syndrome. Sleep Breath 15:679-86
12. Álvarez D, Gutiérrez GC, Marcos JV, et al. (2010) Spectral Analysis of Single-Channel Airflow and Oxygen Saturation Recordings in Obstructive Sleep Apnea Detection, Proc. of the 32nd Ann. Inter. Conf. of the IEEE EMBS, Buenos Aires, Argentina, 2010, pp 847–850
13. Gutiérrez-Tobal GC, Hornero R, Álvarez D, et al. (2012) Linear and nonlinear analysis of airflow recordings to help in sleep apnoea-hypopnoea syndrome diagnosys. Physiol Meas 33:1261-75
14. Witten IH, Frank E, Hall MA (2011) Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann/Elsevier, Burlington, MA
15. Marcos JV, Hornero R, Álvarez D, et al.(2009) Assessment of four statistical pattern recognition techniques to assist in obstructive sleep apnoea diagnosis from nocturnal oximetry, Med Eng Phys 31:971–78
16. Welch PD (1967) The use of fast Fourier transform of the estimation of power spectra: a method based on time averaging over short, modified periodograms. IEEE Trans Aud Elec Acoust AU-15 70–3
17. Bishop CM (2006) Pattern Recognition and Machine Learning. Springer, New York: NY.
18. Freund Y, and Schapire RE (1996) Experiments with a new boosting algorithm. In L. Saitta (Ed.), Thirteenth International Conference on MachineLearning, pp. 148–156. Morgan Kaufmann.

Author: Daniel Álvarez González
Institute: ETSI Telecomunicación, Universidad de Valladolid
Street: Paseo Belén, 15
City: Valladolid
Country: Spain
Email: dalvgon@gmail.com