Automated Analysis of Unattended Portable Oximetry by means of Bayesian Neural Networks to Assist in the Diagnosis of Sleep Apnea

D. Álvarez[1,2], G. C. Gutiérrez-Tobal[2], F. Vaquerizo-Villar[2], V. Barroso-García[2], A. Crespo[1], C. A. Arroyo[1], F. Del Campo[1,2], R. Hornero[2]
[1]Sleep Unit, Río Hortega University Hospital, Valladolid, Spain
[2]Biomedical Engineering Group, University of Valladolid, Valladolid, Spain
Email: daniel.alvarez@gib.tel.uva.es

Abstract — **Sleep apnea-hypopnea syndrome (SAHS) is a chronic sleep-related breathing disorder, which is currently considered a major health problem. In-lab nocturnal polysomnography (NPSG) is the gold standard diagnostic technique though it is complex and relatively unavailable. On the other hand, the analysis of blood oxygen saturation (SpO_2) from nocturnal pulse oximetry (NPO) is a simple, noninvasive, highly available and effective alternative. This study focused on the design and assessment of a neural network (NN) aimed at detecting SAHS using information from at-home unsupervised portable SpO_2 recordings. A Bayesian multilayer perceptron NN (MLP-NN) was proposed, fed with complementary oximetric features properly selected. A dataset composed of 320 unattended SpO_2 recordings was analyzed (60% for training and 40% for validation). The proposed Bayesian MLP-NN achieved 94.2% sensitivity, 69.6% specificity, and 89.8% accuracy in the test set. Our results suggest that automated analysis of at-home portable NPO recordings by means of Bayesian MLP-NN could be an effective and highly available technique in the context of SAHS diagnosis.**

Keywords — **Automated diagnosis, Bayesian neural networks, oximetry, regression, sleep apnea-hypopnea syndrome, unsupervised portable monitoring.**

I. Introduction

Sleep apnea-hypopnea syndrome (SAHS) has become a major health problem in industrialized countries due to its chronic nature and relatively high prevalence [1]. SAHS is characterized by recurrent cessations of breathing due to complete or partial collapse of the upper airway. Repetition of apneic events throughout the night prevent patients from resting while sleeping [1]. Nevertheless, SAHS is considered underdiagnosed [2].

In-laboratory nocturnal polysomnography (NPSG) is the gold standard diagnostic method for SAHS [3]. A thoroughly analysis of a NPSG study may last up to 3 hours [4]. Regarding NPSG, two major limitations arise, which contribute to collapse current facilities [5]: complexity of recording and analysis procedures and availability of specialized sleep units. Thus, simplified alternatives are essential to improve SAHS management.

Nocturnal pulse oximetry (NPO) is a simple and noninvasive technique widely used to analyze the ventilatory process [6]. Previous studies demonstrated the usefulness of peripheral blood oxygen saturation (SpO_2) from NPO in the context of SAHS diagnosis [7, 8]. Currently, there exist several commercial portable pulse oximeters, which increases its availability in order to speed up diagnosis. Therefore, the analysis of SpO_2 from NPO has been proposed as a promising methodology to assist in SAHS diagnosis. Nevertheless, most of this research has been carried out in a controlled sleep laboratory. In the present study, we analyze the diagnostic performance of portable unsupervised oximetry carried out at patients' home in the context of SAHS diagnosis.

We hypothesized that automated analysis can enhance the diagnostic ability of unsupervised portable NPO. Previous studies assessed different pattern recognition techniques in the context of SAHS detection, such as linear discriminant analysis, logistic regression, support vector machines (SVM), and neural networks (NN) [9-11]. NN have been previously applied both for classification [12] and regression [13] purposes. In this study, we proposed a thoroughly methodology for automated signal processing based on 3 stages: (i) feature extraction, by means of statistical moments, spectral analysis and nonlinear methods; (ii) feature selection, by means of the fast correlation-based filter (FCBF); and (iii) pattern recognition, using a binary multilayer perceptron NN (MLP-NN) with Bayesian training to classify subject suspected of suffering from SAHS. The aim of this study is twofold: first, to design and assess a classification MLP-NN with high generalization ability and (ii) to validate unsupervised portable NPO at patients' home in the context of SAHS diagnosis.

II. Subjects and Signals

All subjects involved in the study underwent complete in-lab NPSG (E-series, Compumedics) at the Sleep Unit of the Río Hortega University Hospital in Valladolid (Spain) due to high clinical pre-test probability of suffering from moderate-to-severe SAHS. The American Academy of Sleep Medicine (AASM) scoring rules were applied to quantify the apnea-hypopnea index (AHI), which was used to confirm or exclude the disease [14]. An AHI $\geq$ 10 events

per hour of sleep (e/h) were considered as SAHS-positive. In addition, all subjects carried out unsupervised portable NPO at home. A WristOx2 3150 portable pulse oximeter (Nonin) was used to carry out unsupervised NPO. SpO_2 was recorded and stored at a sampling rate of 1 Hz and subsequently analyzed offline. In-lab and at-home sleep studies were accomplished in consecutive nights in order to reduce night-to-night variability. All patients received oral plus written information in order to correctly use the device. The Ethical Committee validated the information provided to the patients and approved the protocol. All subjects signed an informed consent to participate in the study.

A total of 320 subjects compose the population under study. The whole dataset was divided into a training set (first 193 consecutive patients, 60%), for optimization and training purposes, and a test set (127 consecutive remaining subjects, 40%), for validation in an independent test set. Table I shows the sociodemographic and clinical characteristics of the population under study.

TABLE I
DEMOGRAPHIC AND CLINICAL CHARACTERISTICS OF THE POPULATION

Characteristics	Whole dataset	Training set	Test set
Subjects (n)	320	193	127
Age (years)	54.8 ± 13.5	54.2 ± 12.8	55.6 ± 14.4
Males (%)	74.1	76.7	70.1
BMI (kg/m^2)	29.2 ± 5.5	29.3 ± 5.4	29.1 ± 5.5
AHI (e/h)	39.2 ± 29.4	38.9 ± 28.7	39.6 ± 30.6

BMI: body mass index; AHI: apnea-hypopnea index

III. METHODOLOGY

Automated feature extraction, selection, and classification stages were applied to unsupervised SpO_2 recordings. Training and test datasets were used to design and assess the proposed methodology, respectively.

A. Feature extraction

A total of 16 features composed each input pattern to the Bayesian MLP-NN. Features were arranged into four complementary feature subsets [10]:

1) *Statistical moments in the time domain*. Mean (*M1t*), variance (*M2t*), skewness (*M3t*), and kurtosis (*M4t*) from the data histogram of amplitudes in the time domain were computed to quantify central tendency, dispersion, asymmetry, and peakedness [15].

2) *Statistics in the frequency domain*. Statistical moments *M1f-M4f* were computed to characterize the histogram of amplitudes from the power spectral density (PSD) function [15]. In addition, the median frequency (*MF*) and the spectral entropy (*SE*) were also computed to characterize the power distribution (degree of flatness).

3) *Conventional spectral measures*. Conventional measures based on total signal power (*PT*), as well as relative power (*PR*) and peak amplitude (*PA*) in the apnea-related frequency band (0.014-0.033 Hz) were computed to analyze the periodicity of desaturations due to apneic events [6, 10].

4) *Nonlinear measures*. Sample entropy (*SampEn*), central tendency measure (*CTM*), and Lempel-Ziv complexity (LZC) were computed to quantify regularity, variability, and complexity of SpO_2 [10].

B. Feature selection

FCBF is a dimensionality reduction technique that selects relevant and non-redundant variables independently of the classification stage. It is based on a normalization of the information gain (*IG*) called the symmetric uncertainty (*SU*) [16]. *SU* is computed as follows:

$$SU_i(X_i,Y) = 2\frac{IG_i(X_i,Y)}{H_i(X_i)+H(Y)}, i = 1,...,p\ , \qquad (1)$$

where X_i represents every single feature involved in the study, Y is the AHI in the context of SAHS diagnosis, and H is the Shannon entropy. Firstly, FCBF ranks features according to their relevance and a threshold is applied to discard irrelevant features (the higher SU_i the more relevant feature). In this study, the *log criterion* was applied [16]. In the second step, a redundancy analysis is accomplished to remove redundant features: $SU_{i,j(featurei,featurej)}$ between each pair of remaining ranked features ($SU_i \geq SU_j$) is computed, so that feature j is removed if $SU_{i,j} \geq SU_i$.

C. Feature classification

This stage is aimed at classifying oximetric patterns composed of selected optimum features into two mutually exclusive classes: SAHS-negative or SAHS-positive. To accomplish this task, we proposed to use MLP-NNs, which are able to stablish complex nonlinear decision boundaries to discriminate classes without assuming any *a priori* statistical distribution of the input data [17, 18]. The maximum likelihood criterion is the conventional technique to optimize NNs, in order to minimize the error function [17]. Nevertheless, in this study, we propose to apply the Bayesian inference approach, which models the posterior probability density function of the weight vector $p(\mathbf{w}|D)$ as follows [19, 20]:

$$p(\mathbf{w}\,|\,D) = \frac{p(D\,|\,\mathbf{w})\,p(\mathbf{W})}{p(D)}, \qquad (2)$$

where **w** is the vector of weights for a particular training set D, $p(\mathbf{w})$ is the prior probability function over weight space, $p(D|\mathbf{w})$ is the likelihood of the training data, and $p(D)$ is a normalization factor. Modeling the probability this way allows better generalization ability than determining an optimum set of weights for a particular training set. Then, the distribution of output values for the MLP-NN can be inferred according to the following expression:

$$p(t \mid \mathbf{x}, D) = \int p(t \mid \mathbf{x}, \mathbf{w})\, p(\mathbf{w} \mid D) dw, \qquad (3)$$

which can be interpreted as the probability of membership of an input pattern **x** to target t, i.e. to the positive class (t=C_1) or to the negative class (t=C_2) [19, 20]. In order to obtain the optimum number of neurons in the hidden layer (N_H), leave-one-out cross-validation (loo-cv) was applied in the training set.

D. Statistical analysis

Kolmogorov-Smirnoff and Levene tests were used to assess normality and homoscedasticity of every single feature. The nonparametric Mann-Whitney U test was applied to search for statistical significant differences (p-value < 0.01). Classification performance was assessed in terms of sensitivity (Se), specificity (Sp), positive (PPV) and negative (NPV) predictive values, positive (LR+) and negative (LR-) likelihood ratios, and accuracy (Acc), which were all derived from a two-class confusion matrix. In addition, the *kappa* coefficient was also computed.

III. Results

A. Training set

The proposed 16 features were computed in the training set to compose the initial feature space. All features reached significant statistical differences (p-value<0.01) between SAHS-negative and SAHS positive groups. Nevertheless, just the following features were automatically selected by FCBF due to its higher relevancy and complementarity: *M1t*, *M3t*, *M4t*, *SE*, *PR*, *SampEn*, *CTM* y *LZC*. Using this optimum feature subset, the Bayesian MLP-NN was optimized in the training set. As can be seen in Fig. 1, N_H=18 was the optimum number of neurons.

B. Test set

The Bayesian MLP-NN was further validated in the independent test set. Table II shows the confusion matrix for the binary classification of patients into SAHS-negative and SAHS-positive classes. The proposed Bayesian MLP-NN reached 94.2% Se, 69.6% Sp, 93.3% PPV, 72.7% NPV, 3.10 LR+, and 0.08 LR- in the test set. Overall Acc and *kappa* in the test set were 89.8% and 0.65, respectively.

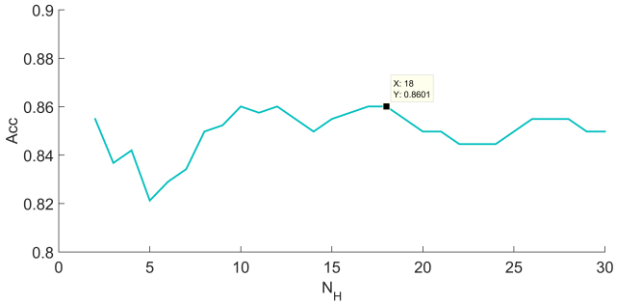

Fig. 1. Optimization of the Bayesian MLP-NN in the training set. Optimum N_H is selected to maximize overall Acc from the loo-cv process in the training set.

TABLE II
CONFUSION MATRIX IN A BINARY CLASSIFICATION CONTEXT FOR THE PROPOSED BAYESIAN MLP-NN IN THE TEST SET

Reference \ Automated		Bayesian MLP-NN	
		SAH- negative	SAHS-positive
PSG	SAHS-negative	**16**	7
	SAHS-positive	6	**98**

IV. Discussion

In this study, a novel approach for automated SAHS detection based on Bayesian MLP-NN was proposed. The methodology was designed and assessed using single-channel SpO_2 recordings from unattended portable NPO at-home. Feature extraction, selection, and classification stages were implemented. A total of 8 features from complementary signal processing techniques were automatically selected and fed the Bayesian MLP-NN. The proposed classifier reached high performance (89.8% Acc) with a slightly unbalanced Se-Sp pair (94.2% *vs.* 69.6%).

Previous studies demonstrated the usefulness of different kinds of NN in the context of SAHS diagnosis, both for classification and regression. In the study by Marcos *et al.*, a conventional MLP-NN fed with nonlinear features from in-lab NPO reached 89.8% Se, 79.4% Sp, and 85.5% Acc in a test set [12]. Similarly, in a subsequent study, Marcos *et al.* assessed a Bayesian MLP-NN fed with 14 features from NPO in the same context, reaching 87.8% Se, 82.4% Sp, and 85.6% Acc in a test set [20]. In a recent study carried out by the same authors, a regression MLP-NN achieved 89.6% Se, 81.3% Sp, and 86.8% Acc in the test set applying the same cutoff for SAHS (AHI=10 e/h) [13]. SVMs have been also applied for SAHS detection from oximetry recordings. In a recent study aimed at detecting moderate-to-severe SAHS, Hang *et al.* designed a SVM-

based classifier using conventional oximetric indexes and achieved 88% Se, 86% Sp, and 87% Acc [21]. Similarly, Álvarez *et al.* used 7 oximetric features automatically selected as inputs to a SVM classifier, reaching 84.2% Acc (84.6% Se and 83.3% Sp) and 84.5% Acc (95.2% Se and 80.0% Sp) in different test sets for a cutoff equal to 10 e/h [10]. In the study by Morillo *et al.*, different classifiers were combined, including NNs and SVMs, in a multiclass approach, achieving 86% Acc and 0.8 *kappa* [11]. Nevertheless, all the studies were carried out in a controlled and supervised sleep laboratory. On the other hand, our approach assessed the diagnostic ability of SpO_2 recordings acquired using a portable pulse oximeter at patient's home.

Some limitations must be taken into account. Firstly, the prevalence of SAHS in the population under study was high (80.3%). Moreover, 56.3% presented severe SAHS (AHI ≥ 30 e/h). This could influence the training process of the NN and, thus, the results of the study (Se-Sp imbalance). Another limitation is related to the recording protocol. The reference in-lab NPSG and portable NPO at patients' home were carried out in different nights. Thus, night-to-night variability could also influence the results.

V. Conclusion

Our results suggest that portable SpO_2 recordings retain high diagnostic ability in an unattended setting. Our approach based on a Bayesian MLP-NN reached high performance in an independent test set using an optimal feature subset from oximetry automatically selected. Therefore, we can conclude that automated analysis of portable at-home SpO_2 recordings by means of Bayesian MLP-NN could be a useful tool to develop an out-of-center screening test for SAHS.

Acknowledgments

This research has been partially supported by Consejería de Educación de la Junta de Castilla y León under project VA059U13 and Pneumology and Thoracic Surgery Spanish Society (SEPAR) under project 265/2012.

References

[1] T. Young, J. Skatrud, and P. E. Peppard, "Risk Factors for obstructive sleep apnea in adults," *JAMA*, vol. 291, pp. 2013-2016, 2009.

[2] S. P. Patil, H. Schneider, A. R. Schwartz, and P. L. Smith, "Adult obstructive sleep apnea. Pathophysiology and diagnosis," *Chest*, vol. 132, pp. 325-337, 2007.

[3] W. W. Flemons, M. R. Littner, J. A. Rowlet, et al., "Home diagnosis of sleep apnea: A systematic review of the literature," *Chest*, vol. 124, pp. 1543-1579, 2003.

[4] J. Fischer, Z. Dogas, C. L. Bassetti, et al., "Standard procedures for adults in accredited sleep medicine centers in Europe," *Journal of Sleep Research*, vol. 21, pp. 357-368, 2012.

[5] T. Gharibeh and R. Mehra, "Obstructive sleep apnea syndrome: natural history, diagnosis and emerging treatment options," *Nature and Science of Sleep*, vol. 2, pp. 233-255, 2010.

[6] C. Zamarrón, P. V. Romero, J. R. Rodríguez, and F. Gude, "Oximetry spectral analysis in the diagnosis of obstructive sleep apnoea," *Clinical Science*, vol. 97, pp. 467-473, 1999.

[7] N. Netzer, A. H. Eliasson, C. Netzer, and D. A. Kristo, "Overnight pulse oximetry for sleep-disordered breathing in adults," *Chest*, vol. 120, pp. 625-633, 2001.

[8] N. A. Collop, S. L. Tracy, V. Kapur, et al., "Obstructive sleep apnea devices for out-of-center (OOC) testing: technology evaluation," *Journal of Clinical Sleep Medicine*, vol. 7, pp. 531-548, 2011.

[9] J. V. Marcos, R. Hornero, D. Alvarez, et al., "Assessment of four statistical pattern recognition techniques to assist in obstructive sleep apnoea diagnosis from nocturnal oximetry," *Med. Eng. Phys.*, vol. 31, pp. 971–978, 2009.

[10] D. Alvarez, R. Hornero, J. V. Marcos, N. Wessel, T. Penzel, M. Glos, and Félix del Campo, "Assessment of feature selection and classification approaches to enhance information from overnight oximetry in the context of sleep apnea diagnosis," *Int. J. Neural Syst.*, vol. 23, pp. 1-18, 2013.

[11] D. Sánchez-Morillo, M. A. López-Gordo, and A. León, "Novel multiclass classification for home-based diagnosis of sleep apnea hypopnea syndrome," *Expert Systems with Applications*, vol. pp. 1654–1662, 2014.

[12] J. V. Marcos, R. Hornero, D. Alvarez, et al., "Utility of multilayer perceptron neural network classifiers in the diagnosis of the obstructive sleep apnea syndrome form nocturnal oximetry," *Computers Methods and programs in Biomedicine*, vol 92, pp. 79-89, 2008.

[13] J. V. Marcos, R. Hornero, D. Alvarez, et al., "Automated prediction of the apnea hipopnea index from nocturnal oximetry recordings," *IEEE Transactions on Biomedical Engineering*, vol. 59, pp. 141-149, 2012.

[14] C. Iber, S. Ancoli-Israel, A. Chesson, et al. *The AASM manual for the scoring of sleep and associated events-rules, terminology and technical specifications*.1st edition. Westchester (IL): American Academy of Sleep Medicine, 2007.

[15] J. D. Jobson, *Applied multivariate data analysis*. New York, Springer-Verlag. New York, 1991.

[16] L. Yu and H. Liu, "Efficient feature selection via analysis of relevance and redundancy," *J. Mach. Learn. Res.*, vol. 5, pp. 1205–1224, 2004.

[17] C. M. Bishop, *Neural Networks for Pattern Recognition* (Oxford: Oxford University Press), 1995.

[18] G. P. Zhang, "Neural networks for classification: a survey," *IEEE Trans. Syst. Man Cybern.*, vol. 30, pp. 451–462, 2000.

[19] I. T. Nabney, *Netlab: Algorithms for Pattern Recognition* (London: Springer), 2002.

[20] J. V. Marcos, R. Hornero, D. Álvarez, I. T. Nabney, F. del Campo, and C. Zamarrón, "The classification of oximetry signals using Bayesian neural networks to assist in the detection of obstructive sleep apnoea syndrome," *Physiol. Meas.*, vol. 31, pp. 375–394, 2010.

[21] L-W. Hang, H-L. Wang, J-H. Chen, et al., "Validation of overnight oximetry to diagnose patients with moderate to severe obstructive sleep apnea," *BMC Pulmonary Medicine*, vol. 15:24, pp. 1-13, 2015.