Cite this document as:

Marcos, J.V., Hornero, R., Álvarez, D., Del Campo, F., Zamarrón, C. and López, M., 2008. Utility of multilayer perceptron neural network classifiers in the diagnosis of the obstructive sleep apnoea syndrome from nocturnal oximetry. *Computer Methods and Programs in Biomedicine*, *92*(1), pp.79-89.

DOI: https://doi.org/10.1016/j.cmpb.2008.05.006

Utility of multilayer perceptron neural network classifiers in the diagnosis of the obstructive sleep apnoea syndrome from nocturnal oximetry

J. Víctor Marcos[1], Roberto Hornero[1], Daniel Álvarez[1], Félix del Campo[2], Carlos Zamarrón[3], Miguel López[1]

[1]Biomedical Engineering Group, E.T.S. Ingenieros de Telecomunicación, University of Valladolid, Spain

[2]Servicio de Neumología, Hospital del Río Hortega, Valladolid, Spain

[3]Servicio de Neumología, Hospital Clínico Universitario, Santiago de Compostela, Spain

AUTHOR'S ADDRESS: J. Víctor Marcos

E.T.S. Ingenieros de Telecomunicación

University of Valladolid

Camino del Cementerio s/n

47011 - Valladolid (Spain)

Phone: +34 983 423000, ext. 5589

Fax: +34 983 423667

E-mail: jvmarcos@gmail.es

Abstract

The aim of this study is to assess the ability of multilayer perceptron (MLP) neural networks as an assistant tool in the diagnosis of the obstructive sleep apnoea syndrome (OSAS). Non-linear features from nocturnal oxygen saturation (SaO_2) recordings were used to discriminate between OSAS positive and negative patients. A total of 187 subjects suspected of suffering from OSAS (111 with a positive diagnosis of OSAS and 76 with a negative diagnosis of OSAS) took part in the study. The initial population was divided into training, validation and test sets for deriving and testing our neural network classifier. Three methods were applied to extract non-linear features from SaO_2 signals: approximate entropy (ApEn), central tendency measure (CTM) and Lempel-Ziv complexity (LZC). The selected MLP-based classifier provided a diagnostic accuracy of 85.5% (89.8% sensitivity and 79.4% specificity). Our neural network algorithm could represent a useful technique for OSAS detection. It could contribute to reduce the demand for polysomnographic studies in OSAS screening.

Keywords: Obstructive sleep apnoea syndrome; nocturnal oxygen saturation; multilayer perceptron neural network; pattern classification; non-linear analysis

1. Introduction

Obstructive sleep apnoea syndrome (OSAS) is characterized by recurrent episodes of complete or partial upper airway occlusion during sleep [1, 2]. Hypoxaemia and hypercapnia associated with apnoeic events lead to frequent arousals, sleep disruption and excessive daytime sleepiness. OSAS adversely affects the cardiovascular system. Long-term effects include hypertension, heart failure and cardiac arrhythmias [2, 3]. Additionally, unrecognized OSAS is reported to be a cause of mortality from traffic and industrial accidents [1].

Epidemiologic studies estimate that OSAS affects 1 to 5% of adult men in western countries [4]. Significant differences have been found for the prevalence of OSAS between male and female populations. An overall ratio of prevalence for men to women of 3.3 has been reported [5]. Moreover, the percentage of undiagnosed patients is estimated at 82% in men and 93% in women [6]. Thus, OSAS can be considered an important concern for public health [2]. Nowadays, nocturnal polysomnography (PSG) is the gold-standard method in OSAS diagnosis [7]. However, PSG presents some drawbacks: it is expensive, complex and time consuming [8]. In addition, the reception capacities of currently existing sleep laboratories are overwhelmed by a growing demand [9]. As a result, simplified diagnostic techniques are desirable.

Nocturnal pulse oximetry represents an alternative to PSG since it is readily available, relatively inexpensive and can be performed at home [8]. Pulse oximetry is a widely used technique in pulmonary medicine, critical care and anaesthesia [10]. It allows to monitor arterial oxygen saturation (SaO_2) recordings in a noninvasively manner. In sleep medicine, nocturnal pulse oximetry provides a useful tool for the analysis of respiratory patterns [11]. Specifically, SaO_2 recordings can be used to detect OSAS. Apnoeic events occurred during sleep are frequently accompanied by oxygen desaturations [10]. As a result, oximetry signals from OSAS patients are characterized by an increased instability, whereas signals

from control subjects tend to present a constant value round 97% of saturation [11]. The utility of SaO_2 in OSAS diagnosis has been analyzed in previous researches [11-19]. Visual inspection of oximetry recordings is the most common form of analysis [12]. Classic oximetry indices computed from nocturnal oximetry represented an important advance for automated OSAS diagnosis. The following indices have been applied: the oxygen desaturation over 2% index (ODI2), 3% index (ODI3) and 4% index (ODI4), and the cumulative time spent below 90% of saturation (CT90) [11, 13, 14]. Signal processing techniques have been also applied to SaO_2 study. Non-linear analysis of this recording has reported promising results in OSAS detection [15-18]. Moreover, the diagnostic capability of information from SaO_2 together with clinical variables has been evaluated [19].

In this study, we propose a neural network-based algorithm to help in OSAS diagnosis from nocturnal oximetry data. Neural networks arise as a powerful tool for classification problems [20, 21]. Neural networks suitably adapt to medical diagnosis since it can be considered a pattern classification task [22]. Several types of network models can be considered for classification [21]. The multilayer perceptron (MLP) was the model selected for our algorithm. Classifiers based on MLP networks are the most widely studied and used neural network. They have been commonly applied in diverse fields such as handwritten recognition, speech recognition, fault detection, bankruptcy prediction or medical diagnosis [20, 21]. In contrast to traditional statistical procedures, MLP neural networks have the ability to learn and generalize without the requirement of any previous knowledge on the underlying data distribution [20, 21, 23].

Previous works have been focused on the assessment of neural networks in the OSAS diagnosis problem. Only clinical characteristics were used in [24, 25]. To develop our neural network algorithm, we suggest applying non-linear features from SaO_2 recordings. The three following non-linear methods were applied to SaO_2 analysis: approximate entropy (ApEn)

[26], central tendency measure (CTM) [27] and Lempel-Ziv complexity (LZC) [28]. Earlier researches have proved their utility in SaO_2 analysis for OSAS detection [15-18].

We hypothesize that neural network processing of these features can provide high diagnostic accuracy in classification of suspected OSAS subjects. The aim of this study is to assess the ability of MLP neural networks as an assistant tool in OSAS diagnosis by using non-linear information from SaO_2.

2. Subjects and signals

Available data were acquired from a total of 187 subjects suspected of suffering from OSAS. Sleep study was carried out from midnight to 8:00 AM in the Sleep Unit of Hospital Clínico Universitario de Santiago de Compostela, Spain. The Review Board on Human Studies at this institution approved the protocol and all subjects gave their informed consent to participate in the study. A Criticare 504 oximeter (CSI, Waukeska, WI, U.S.A.) with a sample frequency of 0.2 Hz was used to record SaO_2 signals. Simultaneously, PSG was performed by using a polygraph (Ultrasom Network, Nicolet, Madison, W.I., U.S.A.). Recordings from PSG were analyzed according to the system of Rechtschaffen and Kales in order to obtain a final diagnosis for each subject [29]. Apnoea was defined as a cessation of airflow for 10 seconds or longer. Hypopnoea was characterized by a reduction, without complete cessation, in airflow of at least 50%, accompanied by a decrease of more than 4% in the saturation of haemoglobin. The average apnoea-hypopnoea index (AHI) was computed for hourly periods of sleep from apnoeic/hypopnoeic episodes captured in PSG. Finally, a threshold of AHI $\geq$ 10 events/h was established to determine the subjects who suffered from OSAS.

A positive diagnosis of OSAS (AHI $\geq$ 10 events/h) was confirmed in 111 subjects (59.36%), while the remaining 76 subjects presented an AHI below the threshold. Our population of 187 patients was randomly divided into training, validation and test sets to

develop and evaluate our neural network algorithm [21]. We assigned 74 subjects to train our MLP networks. The validation set with 30 subjects was used to perform model selection. Classification results were obtained from the test set with 83 subjects. Table 1 summarizes the demographic and clinical data from all the subjects under study and training, validation and test sets.

INSERT TABLE 1 AROUND HERE

---

3. Methods

Our algorithm was composed of three different stages. In the first one, the non-linear features were computed from the SaO_2 recordings. In the second stage, these features were preprocessed in order to properly present them to the classifier. Finally, the neural network-based classification algorithm was applied in the third stage. Network output provided a diagnosis from the input feature vector corresponding to a patient. A scheme of the algorithm is presented in Fig. 1.

INSERT FIGURE 1 AROUND HERE

3.1. Feature extraction

Nocturnal oximetry data were used to classify patients. Oximetry provides useful information to be applied in OSAS detection. It reflects respiratory dynamics during sleep. We applied non-linear analysis from SaO_2 to extract information related to OSAS. The following methods were used: approximate entropy (ApEn), central tendency measure (CTM) and Lempel-Ziv complexity (LZC). Previous studies have proved their utility in OSAS diagnosis by means of SaO_2 processing [15-18]. These methods provide a measure of

regularity, variability and complexity of a time series [26-28]. SaO_2 signals tend to remain constant round 97% in normal subjects, varying slightly with age and independently of ethnicity, gender or weight [11]. Significant changes can be found in patients affected by OSAS because of the recurrent episodes of apnoea, which are frequently accompanied by oxygen desaturations [11]. Therefore, an irregular behaviour can be appreciated in SaO_2 recordings from these patients. It is reflected in regularity, variability and complexity of oximetry signals. As a result, the analysis of these features by means of ApEn, CTM and LZC has reported significant differences between OSAS positive and negative subjects [15-18]. In the present study, we propose to combine information provided by these methods to classify patients suspected of suffering from OSAS. Each oximetry recording was firstly scanned to remove artefacts and drops to zero due to poor contact from the finger probe, patient movements or bad regional circulation. Zero valued samples were excluded from the signal. To estimate ApEn, CTM and LZC, we divided SaO_2 signals into epochs of 200 samples (1000 s, about 17 min). We determined this epoch length according to the apnoea duration (between 25 s and 2 min) in order to include various disordered respiratory events in every epoch [15]. Moreover, an excessively large epoch length would not be effective because of the computational time. The measures obtained from every epoch were averaged to compute final values. Main properties of these methods and user dependant parameters applied for each of them are described below.

3.1.1. Approximate entropy (ApEn)

ApEn is a family of statistics introduced as a quantification of regularity in the data without any a priori knowledge about the system generating them. It was constructed along thematically similar lines to Kolmogorov-Sinai entropy [26]. Several properties of ApEn facilitate its utility for empirical biomedical time series analysis [18, 26]. ApEn is nearly unaffected by noise below a specified filter level, it is scale invariant and model independent.

ApEn evaluates both dominant and subordinated patterns in data. Moreover, it is finite for stochastic, noisy deterministic and composite processes; and increasing values of ApEn correspond to more irregularity in the time series. It discriminates series for which clear feature recognition is difficult. ApEn may correlate with hidden or subclinical changes often undetected by other more classic time series analysis, including moment statistics, spectral analysis and correlation analysis [26]. To perform SaO_2 analysis with ApEn, we set the run length parameter (m) and the size of the tolerance window (r) to 1 and 0.25 times the standard deviation of the original sequence, respectively [18]. It was found that high values of ApEn corresponded to SaO_2 recordings from OSAS positive patients. The presence of OSAS resulted in increased irregularity of oximetry data.

3.1.2. Central Tendency Measure (CTM)

We have used CTM to quantify the degree of variability in SaO_2 signals [27]. A second-order difference plot was generated for each signal to compute CTM [17]. This method evaluates the degree of variability or chaos of a time series rather than defining it as chaotic or not chaotic [27]. A radius (ρ) equal to 0.25 was selected to compute CTM from our SaO_2 signals [15]. The recurrence of apnoeic events in patients with OSAS produced low values of CTM, i.e. oximetry signals from OSAS positive patients presented higher variability than signals from OSAS negative subjects.

3.1.3. Lempel-Ziv Complexity (LZC)

LZC is a non-parametric, simple-to-calculate measure of complexity in a one-dimensional signal. It is related to the number of distinct substrings and the rate of their recurrence along the given sequence [28]. The signal must be transformed into a finite symbol sequence before calculating the complexity measure. The transformation is carried out by comparing each sample with a fixed threshold. We converted SaO_2 signal into a 0-1 sequence to compute LZC. The median value of SaO_2 samples was used as threshold [15]. High values

of LZC were obtained for oximetry signals from OSAS positive subjects, reflecting a high level of complexity.

3.2. Data preprocessing

A total of three parameters composed the input pattern to our neural classifiers. Linear preprocessing of the input features was applied to avoid significant differences between their magnitudes. Features were normalized to have zero mean and unit standard deviation [21]. The re-scaled variables were given by:

$$x_i^n = \frac{\tilde{x}_i^n - \bar{x}_i}{\sigma_i} \tag{1}$$

where n and i are the sample and feature indices, respectively, x_i^n is the normalized value of feature i for sample n, $\tilde{x}_i^n$ is its corresponding raw value, $\bar{x}_i$ is the mean value of feature i and σ_i is its standard deviation.

3.3. Multilayer perceptron neural network classifier

The normalized feature vector was fed to a MLP network for the classification of OSAS positive and negative patients. Generally, neural networks are applied in pattern recognition tasks such as classification and regression. In a classification context, the output of a neural network classifier provides the class membership for the input feature vector. Some advantages can be found on using neural networks for classification problems. First, no prior assumptions about the distribution of the data are required. Neural network algorithms adjust themselves to the environment by means of the training or learning process [20]. Second, it has been proved that MLP networks are universal approximators. They are able to approximate any function with arbitrarily accuracy [30]. Third, neural networks are non-linear models. Thus, real world complex relationships can be modelled by these algorithms [20].

It has been shown that MLP network-based classifiers have provided significant

results in different fields including medical assistance [20, 22]. MLP networks are capable of classifying patterns in an efficient manner since they present the ability to estimate posterior probabilities [21]. The task of a MLP network classifier is to assign new inputs to one of a number of classes. The described OSAS diagnosis problem corresponds to a pattern classification task. Specifically, it can be modelled as a two-class classification problem. Each subject is associated with a feature vector $\mathbf{x} = (x_1, x_2, \ldots, x_I)$. Then, the neural network classifier must provide a decision about its group membership: OSAS positive or OSAS negative. Bayesian decision theory establishes the rule to make such decision in order to minimize the probability of misclassification [20]. Let C denote the membership variable. It takes the values C_1 and C_0 when an input vector corresponds to an OSAS positive subject and an OSAS negative subject, respectively. The Bayesian classification rule states the following:

$$\text{Decide } C_k \text{ for } \mathbf{x} \text{ if } P(C_k|\mathbf{x}) = \max_{k=0,1} P(C_k|\mathbf{x}) \tag{2}$$

where $P(C_k|\mathbf{x})$ is the posterior probability of group C_k. Therefore, an estimation of the posterior probability function of both groups is required to make appropriate decisions. A suitable design of our MLP networks was carried out in order to obtain such estimation from the network output value. Appropriate choices are required about the following network design issues: the output coding scheme, the form of the error function used in network training and the activation function of neurons in the output layer [21].

A two-class classification problem involves the use of a single output neuron. The output value of this neuron is given by the mapping function $f(\mathbf{x}, \mathbf{w})$ performed by the MLP network, where $\mathbf{x}$ represents the input feature vector and $\mathbf{w}$ the vector containing network weights and biases. For a single hidden layer network it is computed as follows:

$$y = f(\mathbf{x}, \mathbf{w}) = g_l\left\{\sum_{h=1}^{H}\left[w_{ho} g_t\left(\sum_{i=1}^{I} w_{ih} x_i + b_h\right) + b_o \right]\right\} \tag{3}$$

where I is the number of features in the input vector, H is the number of hidden neurons, w_{ho} is the weight connecting hidden neuron h with output neuron o, b_o is the bias associated to output neuron o, w_{ih} is the weight connecting the feature i of the input vector with hidden neuron h, b_h is the bias associated to hidden neuron h, $g_t(\cdot)$ is the activation function for neurons in the hidden layer, $g_l(\cdot)$ is the activation function for the output layer neuron and y is the network output value. In our study, we evaluated MLP networks with a single hidden layer of neurons since networks with this architecture are capable of universal approximation [31]. The cross-entropy error function (E_D) was applied for training our MLP networks since they were used for classification [21]. This function is expressed as follows:

$$E_D = -\sum_{n=1}^{N}\left[t_n \ln y_n + (1 - t_n)\ln(1 - y_n)\right] \tag{4}$$

where t_n is the target or desired value for vector $\mathbf{x}_n$ in the training set and y_n is the output value provided by the network. We considered a target coding scheme with $t = 1$ if the input vector belonged to class C_1 (OSAS positive group) and $t = 0$ if the input vector belonged to class C_0 (OSAS negative group). The logistic activation function was used for the single output neuron of our networks. This function is expressed as:

$$g_l(z) = \frac{1}{1 + e^{-z}} \tag{5}$$

where z is the effective input to the neuron, which is computed as the weighted sum of neuron inputs. The output of this function lies in the range between 0 and 1. Thus, this output value can be interpreted as a probability. Finally, the network output value represents the posterior

probability for class C_1. The Bayesian classification rule expressed in equation (2) can be applied to make decisions about the group membership of an input vector. According to the described design criteria, the following properties can be associated with the output of the MLP classifier:

$$P(C_1|\mathbf{x}) = y \Rightarrow P(C_0|\mathbf{x}) = 1 - y \tag{6}$$

The NETLAB software was used to implement our neural network classifiers [32]. We trained our MLP networks by using a training set with 74 patterns. Subjects in this set were randomly selected from the initial population with 187 patients. As suggested in [21], weights and biases were initially generated from a spherically symmetric Gaussian distribution with a mean equal to zero. The variance of the Gaussian was set to $1/(I+1)$ and $1/(H+1)$ for weights and biases in hidden and output layers, respectively. The hyperbolic tangent activation function was used for neurons in the hidden layer since it has been empirically proved that this function provides a fast convergence of training algorithms [21, 23]. The scaled conjugate gradient (SCG) algorithm was used for training our MLP networks [33]. The aim of training is to find the network with the best generalization capability for the proposed problem. Generalization refers to the ability to classify well new input data. It is achieved when the network mapping function $f(\mathbf{x}, \mathbf{w})$ represents the underlying systematic aspects of the data rather than capturing the specific details in the training set [21]. When the latter occurs, it is said that the network has overfitted the training data. In order to avoid this problem, weight decay regularization was applied [21, 34]. A penalty term (E_W) proportional to the sum of squared weights was added to the cross-entropy error function (E_D). The objective function (E) to be minimized through training was given by:

$$E = E_D + E_W = -\sum_{n=1}^{N}\left[t_n \ln y_n + (1-t_n)\ln(1-y_n)\right] + \alpha \frac{1}{2}\sum_{w \in W} w^2 \tag{7}$$

where W denotes the set of weights in the network excluding biases and α is a regularization parameter [21]. The weight decay function encourages optimization algorithm to produce networks with small weights. The regularization parameter (α) controls the effective complexity of the model [34]. A large value of α will lead to over-smoothing mappings or underfitting whereas a small value of α could produce overfitted mappings [21, 23].

4. Results

The goal of our experiments was to find the network algorithm having the best generalization performance. Generalization is influenced by three factors: the size of the training set and how of representative is of the environment of interest, the architecture of the neural network and the physical complexity of the problem [23]. The two first factors are closely related whereas the latter cannot be controlled. Theoretically, the training set should grow according to the complexity of the network to prevent overfitting and poor generalization [23]. However, the size of the training set is generally fixed for real problems and the matter becomes to find the most appropriate network architecture [30]. In this context, the trade-off between bias and variance arises [20, 21]. A simple or inflexible model can lead to underfit the data. The model may not have the ability to learn enough the underlying distribution. It is said that the neural network classifier has a large bias. On the other hand, a too complex or flexible model could capture the noise present in the training data, leading to overfitting. In this case, the neural network classifier has a large variance. Both situations result in poor generalization and a good trade-off between bias and variance is desired in order to obtain a useful neural network classifier [20].

The architecture of the network, i.e. the number of neurons in the hidden layer, is the remaining decision regarding network design. The number of hidden neurons corresponds to the complexity of the problem. We evaluated the performance of different MLP network

configurations by measuring the accuracy reached on the validation set [21]. The number of hidden neurons was varied from a minimum of 2 up to a maximum of 50 units. Since a particular training run is sensitive to the initial value of the weights, the accuracy on the validation set of each network configuration was measured and averaged for a total of 25 runs. Accuracy quantifies the total number of subjects correctly classified. It is computed as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{8}$$

where false negatives (*FN*) are the OSAS positive patients classified as control subjects, and false positives (*FP*) are the controls classified as patients. True positives (*TP*) and true negatives (*TN*) are the patients and control subjects correctly recognized, respectively. In addition, several values of the regularization parameter (α) were evaluated. The effect of varying the size of the hidden layer and the regularization parameter is depicted in Fig. 2.

INSERT FIGURE 2 AROUND HERE

We found that the highest classification accuracy was reached with a regularization parameter of α = 0.25 (curve plotted with symbol ■). This parameter controls the effective network size and has a direct influence on classification performance. Weight decay generally lowers the generalization error [34]. However, a too high value of the regularization parameter resulted in an excessively simple network classifier. On the other hand, a too small value of this parameter led to network models with high variance [34]. This behaviour can be observed in our results. Classification accuracy increased according to the regularization parameter until it reached a value of 0.25. Larger values of the parameter produced network classifiers with lower classification performance on the validation set. Additionally, the best performance for the MLP networks trained with α = 0.25 was achieved with 10 neurons in the

hidden layer. Increasing the size of the hidden layer resulted in lower values of classification accuracy. Thus, a configuration with 10 hidden neurons and α = 0.25 was chosen. The solution with the highest accuracy on the validation set was selected as the final neural network algorithm from the 25 trained MLP networks with such configuration. The Hinton diagram of this network is showed in Fig. 3. The size of the squares corresponds to the strengths of the connections. White and dark colours mark positive and negative weights, respectively.

INSERT FIGURE 3 AROUND HERE

We assessed the performance of the selected classifier by using the test set with 83 subjects. Sensitivity, specificity and accuracy were computed. Sensitivity reflects the rate of OSAS positive subjects correctly classified while specificity is equal to the true negative rate:

$$Sensitivity = \frac{TP}{TP + FN} \tag{9}$$

$$Specificity = \frac{TN}{TN + FP} \tag{10}$$

Moreover, we applied receiver operating characteristic (ROC) analysis [35, 36]. ROC curve analysis suppresses the requirement for a threshold value by evaluating the performance of the classifier over the whole range of possible values. A plot of sensitivity versus 1-specificity is made over a range of threshold values to obtain the ROC curve. The area under ROC curve (AROC) was used as a measure of classification performance [35]. It represents the probability of correct classification for a randomly chosen pair of OSAS positive and OSAS negative subjects [36].

Several methods can be applied to estimate the classification performance of our

neural network algorithm. For instance, resubstitution, hold-out, cross-validation and bootstrap can be considered [37]. In the present research we computed the results of our MLP network classifiers by means of the hold-out technique. The selected network algorithm was evaluated on the test set with 83 patients. A diagnostic accuracy of 85.5% was provided by the MLP network classifier developed in this study. Sensitivity and specificity were 89.8% and 79.4%, respectively. The classifier yielded an AROC of 0.90. Additionally, in order to validate our results, hold-out testing was done by using bootstrap [38]. Several test sets were generated by bootstrapping to assess the classification capability of the selected MLP network. Our initial test set with 83 subjects was resampled with replacement for a total of 10000 times to produce new test sets of the same size [39]. Sensitivity, specificity, accuracy and AROC were computed. These measures were averaged to calculate the bootstrap estimation. A sensitivity of 90.8%, a specificity of 79.1%, an accuracy of 86.0% and an AROC of 0.91 were obtained. The results computed by bootstrapping were similar to those estimated by using the hold-out method. This reinforces the results reported in the present study.

To assess the improvement provided by the proposed algorithm, the classification performance of our MLP network was compared with that reached by means of classic oximetry indices. The following indices were provided by our oximetry equipment: ODI2, ODI3, ODI4 and CT90. We determined the optimum decision threshold for each of these indices from ROC analysis of training data. The threshold that provided the highest accuracy on the training set was selected as optimum. A threshold value of 9, 8 and 7.6 events/h was established for ODI2, ODI3 and ODI4, respectively, while a percentage of time during sleep of 11% was determined for CT90. Sensitivity, specificity and accuracy for each of the indices were computed by applying these thresholds on data from our test set. In addition, we performed ROC analysis on data from subjects in the test set to compute the AROC value for

each oximetry index. Classification results provided by our proposed MLP algorithm and oximetry indices are summarized in Table 2. ROC curves are displayed in Fig. 4.

INSERT TABLE 2 AROUND HERE

INSERT FIGURE 4 AROUND HERE

The four indices provided high specificity and low sensitivity. The highest classification accuracy was provided by ODI3 with a rate of 73.1%. The best AROC value was obtained by means of ODI4 with 0.80. Both accuracy and AROC achieved by these indices were significantly outperformed by our MLP classifier.

5. Discussion

A neural network-based classification algorithm to assist OSAS diagnosis was presented. Our algorithm only used information from nocturnal oximetry recordings. Non-linear analysis of SaO_2 signals was performed for feature extraction. The following non-linear methods were applied: ApEn, CTM and LZC. A MLP neural network classifier was chosen to process these features in order to discriminate between OSAS positive and negative subjects. Finally, the selected MLP network algorithm provided a diagnostic accuracy of 85.5% (89.8% sensitivity and 79.4% specificity) and an AROC of 0.90. Our algorithm allows to improve sensitivity, specificity and accuracy provided by classic oximetry indices, as it can be observed from results in Table 2.

The Hinton diagram in Fig. 3 supplies useful information about the classification algorithm developed in the present study. The three input features had a relevant influence on the network output value. The magnitude of the weights in the hidden layer associated to each

of these features was comparable and none of them was ignored by the network during training. However, the task performed by some hidden neurons was insignificant, since their weights had a value near zero as a result of regularization. Hence, classification of input patterns was mainly carried out by hidden neurons H3, H6 and H10, together with H5, H7 and H9. The magnitude of weights in the output layer represents the contribution of each hidden neuron in the network. It can be observed that the weights connecting these neurons with the output neuron had a larger magnitude than that of the weights operating on the other hidden neurons. Weight decay regularization penalizes large weights and, thus, rewards simpler mappings [34]. Overtraining was avoided by reducing the effective complexity of the network. Therefore, generalization capability was increased.

Our MLP classifier has provided promising results in OSAS diagnosis problem with classification accuracy of 85.5% and AROC of 0.90. The proposed approach could provide an accurate tool to detect the presence of OSAS in an initial test of the patient. Only nocturnal oximetry recording is required for our neural network-based OSAS diagnosis technique. The sensitivity provided by the MLP network was 89.8%. A total of 5 OSAS positive were misclassified when our algorithm was evaluated on the test set with 83 patients. Two of them presented an AHI of 11 events/h. Other two subjects had an AHI of 19 and 29 events/h, respectively. Only one of these five patients presented severe OSAS, i.e., AHI $\geq$ 30 events/h. The specificity achieved by the MLP network was 79.4%. An incorrect diagnosis of OSAS was made for 7 OSAS negative subjects in the test set. Three of them suffered from chronic obstructive pulmonary disease (COPD) [40]. The values of ApEn and LZC computed from SaO_2 signals of these subjects were substantially greater than the mean values of these features computed from control subjects in the training set. Similarly, their CTM values were lower than the mean CTM value. Therefore, the presence of COPD directly affects the behaviour of SaO_2 recording during sleep. This fact is reflected in the values of ApEn, CTM

and LZC, resulting in incorrect decision of the algorithm.

Sensitivity is the critical parameter in medical diagnosis. The penalty for misclassifying an OSAS positive subject is greater than that for misclassifying an OSAS negative. The classification of the selected MLP network can be made by taking into account this interpretation. Then, we assigned a cost for misclassifying an OSAS positive patient that was twice the cost for misclassifying a normal subject, i.e., the sensitivity of the classifier was prioritized. A sensitivity of 91.8%, a specificity of 67.7% and an accuracy of 81.7% were achieved by the network when misclassification costs were included in the decision rule. Although sensitivity was increased, specificity substantially decreased regarding the application of the Bayes decision rule with equal costs of misclassification for both groups of subjects.

The high prevalence of OSAS and the inconvenience related to PSG has motivated researches to look for new valid alternatives. Nocturnal oximetry signal has been commonly used in previous studies regarding to OSAS diagnosis. Visual inspection of SaO_2 is a time-consuming task. A sensitivity of 91% and a specificity of 69% have been reported by means of this technique on a population of 96 patients [12]. Classic oximetry indices facilitate SaO_2 analysis. Oximeters can provide a measure of ODI2, ODI3, ODI4 and the cumulative time spent below a given level of saturation. Netzer et al. [11] reviewed several studies where the diagnostic capability of these indices was assessed. The reported values of sensitivity ranged from 31 to 98% and for specificity from 41 to 100% [11]. Vazquez et al. [13] reported the highest diagnostic accuracy (98% sensitivity and 88% specificity) by using ODI4. However, a non-standard pulse oximetry test and a conservative definition of OSAS (AHI $\geq$ 15 events/h) were used [41]. Magalang et al. [14] proposed a combination of oximetry indices to predict AHI. A large database with 224 subjects was available to develop the prediction model. It was prospectively validated on two validation sets with 101 and 191 patients, respectively. A

threshold of AHI ≥ 15 events/h was applied to define the presence of OSAS. This study reported a sensitivity of 90% and a specificity of 70%. In the study by Roche et al. [19] information from SaO_2 was combined with clinical data to develop an OSAS diagnosis model. A diagnostic accuracy of 62.1% was obtained. Finally, previous studies analyzed the utility of ApEn, CTM and LZC from SaO_2 data in OSAS diagnosis [15-18]. Significant differences between OSAS negative and positive subjects were found with these non-linear methods. Neural networks provide an effective means to simultaneously process these features in order to achieve high accuracy in OSAS diagnosis. Our MLP classifier outperformed the classification performance reached by means of ApEn, CTM and LZC individually. The studies in [15-18] were developed with the same database of 187 patients used in this research. However, their methodology differs from that applied in the present study. Neural network classifiers were not used in these previous studies. Additionally, other differences can be pointed out. In [15-17], ApEn, CTM and LZC were optimized and tested on a same set of SaO_2 signals. Therefore, the classification results reported in these studies may be optimistically biased since the resubstitution method was used to estimate them [37]. In [18], the initial population was divided into training and test sets. Training was performed with the same set of 74 subjects that has been used in the present research to train our MLP networks. Nevertheless, a different test set was used to compute final results.

Neural network represent a suitable assistant tool for different medical tasks such as diagnosis or outcome prediction [22]. Our proposed method represents a step forward in automated OSAS diagnosis from nocturnal oximetry. Other previous studies have applied neural networks to OSAS diagnosis. MLP neural networks were evaluated using our database [42]. Spectral and non-linear features from SaO_2 signals were used as network inputs. A mean sensitivity of 86.5% and a mean specificity of 81.4% were achieved by the MLP network configuration determined in [42]. The classification algorithm developed here had better

sensitivity (89.8%). However, several differences can be found with regard to the present research. The set of input features fed into the network, the optimization algorithm used for network training and the computation of classification results differ from one study to another. Moreover, a different methodology was applied to design the classifiers since a probabilistic interpretation can be drawn from the output value provided by the MLP algorithm presented here. In [43], RBF network classifiers were analyzed by using our database. ApEn, CTM and LZC composed the network input pattern. The classification results achieved in that study were comparable to those provided by our MLP-based algorithm. A sensitivity of 89.4% and a specificity of 81.4% were obtained by a RBF network trained with the *k*-means algorithm [43]. However, these results were computed by averaging the sensitivity and specificity statistics achieved by different RBF classifiers with the same network architecture. Therefore, a specific classification algorithm was not provided. In [44], three clustering techniques were evaluated by only using 74 subjects from our database. The same dataset was used to optimize the proposed method and to test it. Thus, the results could be biased. On the other hand, only clinical data were used as network inputs by other researches. Kirby et al. [25] proposed a generalized regression neural network. Network input was composed of 23 clinical variables. A sensitivity of 98.9% and a specificity of 80% were reported. El-Solh et al. [24] achieved 94.9% sensitivity and 64.7% specificity using a feedforward neural network trained with the backpropagation algorithm. A total of 12 variables from clinical and anthropomorphic measurements were fed into the network. However, the methodology applied in these studies differs from our method. As suggested in [21], we used a validation set to perform model selection, i.e., to compare trained networks of different size or complexity. The final selected network could include some overfitting on the test set when only training and test sets are used [21].

Methodological differences should be taken into account in order to compare our

results with those reported by the cited studies. Clinical and demographic characteristics of the population under study and its number of samples vary from one research to another. Thus, a public database of SaO_2 recordings would be desirable. Additionally, we noticed that sensitivity was usually higher than specificity for the mentioned OSAS diagnosis methods as well as for our neural network classifier. As indicated before, high sensitivity may be preferable from a clinical point of view since the cost for misclassifying an OSAS positive subject is greater than that for misclassifying an OSAS negative.

The purpose of our future research is to improve the classification results achieved by our MLP algorithm. Additional features can be considered for the input pattern. Combination of linear and non-linear parameters from SaO_2 signals represents an interesting research line. As indicated before, a preliminary study with spectral and non-linear features from oximetry data was developed [42]. Although oximetry indices from our recordings have shown poor classification ability, they could also complement non-linear features to discriminate OSAS patients. Moreover, other neural network classifiers should be assessed.

Some limitations could be found in our study. Even larger database would be advantageous. Firstly, more accurate results could be obtained from the test process. Secondly, a better final solution for the neural network algorithm could be derived from training since the number of samples in the training set influences the generalization capability of a MLP network [21, 23]. However, available data are usually limited in real problems. Then, a significant percentage should be assigned for training whereas an adequate quantity of data is required for other tasks such as model selection or test. The possible noise in SaO_2 signals is another limitation. The artefact rate is high in overnight pulse oximetry due to movements during sleep. Noisy components could artificially increase the measures of regularity, variability and complexity provided by non-linear methods, resulting in classification error. The presence of other breathing disorders in OSAS negative subjects,

such as COPD, leads to a similar consequence. It can be pointed out that COPD in control subjects causes our method to classify them incorrectly, resulting in higher sensitivity than specificity. Finally, the data collection process could be enhanced, since new oximeters can work at higher sampling frequencies. It allows to improve artefact detection due to patient movements [14, 45, 46]. However, the study carried out by Warley et al. [47] showed that a sampling frequency of 0.2 Hz provides reasonable resolution in SaO_2 signals. Although there exists some underestimation of the peak SaO_2 in recovery post-apnoea, the shape of the signals is preserved and an accurate estimation of their regularity, variability and complexity can be obtained [15-18]. In addition, the averaging time of the oximeter should be considered. The oximeter used in our research has a time response equal to the last 16 R-R intervals, which is higher than the averaging time of 3 s (at a heart rate of 80 beats per minute or more) recommended in [48]. It has been proved that the time response affects oxygen desaturation indices [49]. Low sampling rate of our oximetry equipment may have affected the results of ODI calculations in our study. Nevertheless, the aim of our study is not the detection of desaturation events. As it was shown in [15-18], a valid non-linear analysis of our oximetry data can be performed by means of ApEn, CTM and LZC.

In summary, we verified that MLP neural networks were a suitable tool to improve OSAS diagnosis from SaO_2 signals. Non-linear processing methods applied in this study (ApEn, CTM and LZC) provided useful information from SaO_2 recordings in order to classify suspected OSAS patients. The performance of our neural network classifier was measured on a test set with 83 subjects suspected of suffering from OSAS. An accuracy of 85.5% and an AROC of 0.90 were achieved by the developed MLP neural network, improving the performance of visual inspection and classic automated analysis of SaO_2 signals. Therefore, the proposed method could be useful for OSAS detection.

Acknowledgements

This research has been supported by Consejería de Educación de la Junta de Castilla y León under project VA108A06.

References

[1] K.G. Van Houwelingen, R. Van Uffelen and A.C.M. Van Vliet, The sleep apnoea syndromes, European Heart Journal 20 (1999) 858-66.

[2] A. Qureshi and R.D. Ballard, Obstructive sleep apnea, The Journal of Allergy and Clinical Immunology 112 (2003) 643-651.

[3] S.M. Caples, A.S. Gami and V.K. Somers, Obstructive sleep apnea, Annals of Internal Medicine 142 (2005) 187-197.

[4] T. Young, P.E. Peppard and D.J. Gottlieb, Epidemiology of obstructive sleep apnea: a population health perspective, American Journal of Respiratory and Critical Care Medicine 165 (2002) 1217-1239.

[5] E.O. Bixler, A.N. Vgontzas, H.M. Lin, T.T. Have, J. Rein, A. Vela-Bueno and A. Kales, Prevalence of sleep-disordered breathing in women. Effects of gender, American Journal of Respiratory and Critical Care Medicine 163 (2001) 608-613.

[6] T. Young, L. Evans, L. Finn and M. Palta, Estimation of the clinically diagnosed proportion of sleep apnea syndrome in middle-aged men and women, Sleep 20 (1997) 705-706.

[7] Agency for Health Care Policy and Research (AHCPR), Systematic review of the literature regarding the diagnosis of sleep apnea: summary, evidence report/technology assessment: number 1 (Department of Health and Human Services, U.S. Public Health Service, 1999).

[8] J.A. Bennet and W.J.M. Kinnear, Sleep on the cheap: the role of overnight oximetry in the diagnosis of sleep apnoea hypopnoea syndrome, Thorax 11 (1999) 958-959.

[9] F. Gagnadoux, N. Pelletier-Fleury, C. Philippe, D. Rakatonanahary and F. Bernard, Home unattended vs hospital telemonitored polysomnography in suspected obstructive sleep apnea syndrome: a randomized crossover trial, Chest 121 (2002) 753-758.

[10] K.E. Bloch, Getting the most out of nocturnal pulse oximetry, Chest 124 (2003) 1628-1630.

[11] N. Netzer, A.H. Eliasson, C. Netzer and D.A. Kristo, Overnight pulse oximetry for sleep-disordered breathing in adults: a review, Chest 120 (2001) 625-633.

[12] J.M. Rodríguez, P. De Lucas, M.J. Sánchez, J.L. Izquierdo, R. Peraita and J.M. Cubillo, Usefulness of the visual analysis of night oximetry as a screening method in patients with suspected clinical obstructive sleep apnea syndrome, Archivos de Bronconeumología 32 (1996) 437-441.

[13] J.C. Vázquez, W.H. Tsai, W.W. Flemons, A. Masuda, R. Brant, E. Hajduk, W.A. Whitelaw and J.E. Remmers, Automated analysis of digital oximetry in the diagnosis of obstructive sleep apnoea, Thorax 55 (2000) 302-307.

[14] U.J. Magalang, J. Dmochowski, S. Veeramachaneni, A. Draw, M.J. Mador, A. El-Solh and B.J.B. Grant, Prediction of the apnea-hypopnea index from overnight pulse oximetry, Chest 124 (2003) 1694:1701.

[15] D. Álvarez, R. Hornero, D. Abásolo, F. Del Campo and C. Zamarrón, Nonlinear characteristics of blood oxygen saturation from nocturnal oximetry for obstructive sleep apnoea detection, Physiological Measurement 27 (2006) 399-412.

[16] F. Del Campo, R. Hornero, C. Zamarrón, D. Abásolo and D. Álvarez, Oxygen saturation regularity analysis in the diagnosis of obstructive sleep apnea, Artificial Intelligence in Medicine 37 (2006) 111-118.

[17] D. Álvarez, R. Hornero, M. García, F. Del Campo and C. Zamarrón, Improving diagnostic ability of blood oxygen saturation from overnight pulse oximetry in obstructive

sleep apnea detection by means of central tendency measure, Artificial Intelligence in Medicine 41 (2007) 13-24.

[18] R. Hornero, D. Álvarez, D. Abásolo, F. Del Campo and C. Zamarrón, Utility of aproximate entropy from overnight pulse oximetry data in the diagnosis of the obstructive sleep apnea syndrome, IEEE Transactions on Biomedical Engineering 54 (2007) 107-13.

[19] N. Roche, B. Herer, C. Roig and G. Huchon, Prospective testing of two models based on clinical and oximetric variables for prediction of obstructive sleep apnea, Chest 121 (2002) 747-752.

[20] G.P. Zhang, Neural networks for classification: a survey, IEEE Transactions on Systems, Man and Cibernetics, Part C 30 (2000) 451-462.

[21] C.M. Bishop, Neural networks for pattern recognition (Oxford University Press, Oxford UK, 1995).

[22] W.G. Baxt, Application of artificial neural networks to clinical medicine, Lancet 346 (1995), 1135-1138.

[23] S. Haykin, Neural networks: a comprehensive foundation (Prentice Hall Inc., New Jersey NJ, 1999).

[24] A.A. El-Solh, M.J. Mador, E. Ten-Brock, D.W. Shucard, M. Abul-Khoudoud and B.J.B. Grant, Validity of neural network in sleep apnea, Sleep 22 (1999) 105-111.

[25] S.D. Kirby, W. Danter, C.F.P. George, T. Francovic, R.R.F. Ruby and K.A. Ferguson, Neural network prediction of obstructive sleep apnea from clinical criteria, Chest 116 (1999) 409-415.

[26] S.M. Pincus, Assessing serial irregularity and its implications for health, Annals of the New York Academy of Science 954 (2001) 245-267.

[27] M.E. Cohen, D.L. Hudson and P.C. Deedwania, Applying continuous chaotic modeling to cardiac signals, IEEE Engineering in Medicine and Biology Magazine 15 (1996) 97-102.

[28] A. Lempel and J. Ziv, On the complexity of finite sequences, IEEE Transactions on Information Theory 22 (1976) 75-81.

[29] A. Rechtschaffen and A. Kales, A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects (Brain Information Services, Brain Research Institute, University of California, Los Angeles, 1968).

[30] D.R. Hush and B.G. Horne, Progress in supervised neural networks, IEEE Signal Processing Magazine 10 (1993) 8-39.

[31] K. Hornik, Approximation capabilities of multilayer feedforward networks, Neural Networks 4 (1991) 251-257.

[32] I.T. Nabney, NETLAB: Algorithms for pattern recognition (Springer, Berlin, 2002).

[33] M.F. Moller, A scaled conjugate gradient algorithm for fast supervised learning, Neural Networks 6 (1993) 525-533.

[34] J.E. Moody, Note on generalization, regularization, and architecture selection in nonlinear learning systems, Proceedings of the IEEE Workshop in Neural Networks for Signal Processing (1991) 1-10.

[35] A.P. Bradley, The use of the area under the ROC curve in the evaluation of machine learning algorithms, Pattern Recognition 30 (1997) 1145-1159.

[36] J.A. Hanley and B.J. McNeil, The meaning and use of the area under a receiving operating characteristic (ROC) curve, Radiology 143 (1982) 29-36.

[37] A.K. Jain, R.P.W. Duin and J. Mao, Statistical pattern recognition: a review, IEEE Transactions on Pattern Analysis and Machine Intelligence 22 (2000) 4-37.

[38] M.L. Raymer, W.F. Punch, E.D. Goodman, L.A. Kuhn and A.K. Jain, Dimensionality reduction using genetic algorithms, IEEE Transactions on Evolutionary Computation 4 (2000) 164-171.

[39] B. Efron, Bootstrap methods: another look at the jackknife, The Annals of Statistics 7 (1979) 1-26.

[40] R.A. Pauwels, A.S. Buist, P.M.A. Calverley, C.R. Jenkins and S.S. Hurd, Global strategy for the diagnosis management, and prevention of chronic obstructive pulmonary disease, American Journal of Respiratory and Critical care Medicine 163 (2001) 1256-1276.

[41] The Atlas Task Force, EEG arousals: scoring rules and examples. A preliminary report from the Sleep Disorder Task Force of the American Sleep Disorders Association, Sleep 15 (1992) 173-184.

[42] J.V. Marcos, R. Hornero, D. Álvarez, F. Del Campo and M. López, Applying neural network classifiers in the diagnosis of the obstructive sleep apnea syndrome from nocturnal pulse oximetric recordings, Proceedings of the 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (2007) 5174-5177.

[43] J.V. Marcos, R. Hornero, D. Álvarez, F. Del Campo, M. López and C. Zamarrón, Radial basis function classifiers to help in the diagnosis of the obstructive sleep apnoea syndrome from nocturnal oximetry, Medical and Biological Engineering and Computing 46 (2008) 323-332.

[44] D. Álvarez, R. Hornero, J.V. Marcos, F. Del Campo and M. López, Obstructive sleep apnea detection using clustering classification of nonlinear features from nocturnal oximetry, Proceedings of the 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (2007) 1937-1940.

[45] W.W. Flemons, M.R. Littner, J.A. Rowley, P. Gay, W. McDowell, D.W. Hudgel, R. Douglas and D.I. Loube, Home diagnosis of sleep apnea: a systematic review of the literature, Chest 124 (2003) 1543-1579.

[46] P. Levy, J.L. Pepin, C. Deschaux-Blanc, B. Paramelle and C. Brambilla, Accuracy of oximetry for detection of respiratory disturbances in sleep apnea syndrome, Chest 109 (1996) 395-399.

[47] A.R. Warley, J.H. Mitchell and J.R. Stradling, Evaluation of the Ohmeda 3700 pulse oximeter, Thorax 42 (1987) 892-896.

[48] C. Iber, S. Ancoli-Israel, A.L. Chesson and S.F. Quan, The AASM manual for the scoring of sleep and associated events: rules, terminology, and technical specifications (Westchester, Ill: American Academy of Sleep Medicine, 2007).

[49] R. Farré, J.M. Montserrat, M. Rotger, E. Ballester and D. Navajas, Accuracy of thermistors and thermocouples as flow-measuring devices for detecting hypopneas, European Respiratory Journal 11 (1998) 179-182.

Tables

Table 1. Demographic and clinical statistics of all subjects, training set and test set. Data are presented as mean ± standard deviation. OSAS Positive: patients with a positive diagnosis of obstructive sleep apnoea syndrome; OSAS Negative: patients with a negative diagnosis of obstructive sleep apnoea syndrome; BMI: body mass index; AHI: apnoea/hypopnoea index computed as events for hourly periods.

ALL SUBJECTS			
	All	OSAS Positive	OSAS Negative
Subjects (n)	187	111	76
Age (years)	57.97 ± 12.84	58.30 ± 12.88	57.57 ± 12.87
Males (%)	78.61	84.68	69.74
BMI (kg/m^2)	29.54 ± 5.51	30.45 ± 4.92	28.42 ± 6.02
Recording Time (h)	8.19 ± 0.62	8.17 ± 0.75	8.22 ± 0.33
AHI (n/h)		40.07 ± 19.64	2.04 ± 2.36
TRAINING SET			
	All	OSAS Positive	OSAS Negative
Subjects (n)	74	44	30
Age (years)	58.25 ± 12.14	56.73 ± 13.61	59.59 ± 10.19
Males (%)	75.68	79.55	70.00
BMI (kg/m^2)	29.62 ± 5.71	30.19 ± 5.09	28.93 ± 6.40
Recording Time (h)	8.22 ± 0.41	8.20 ± 0.49	8.25 ± 0.27
AHI (n/h)		38.11 ± 18.18	2.60 ± 2.51
VALIDATION SET			
	All	OSAS Positive	OSAS Negative
Subjects (n)	30	18	12
Age (years)	57.42 ± 15.65	59.08 ± 13.47	55.75 ± 18.19
Males (%)	76.67	88.89	58.33
BMI (kg/m^2)	30.92 ± 6.11	32.37 ± 6.05	29.33 ± 6.07
Recording Time (h)	8.05 ± 0.80	7.86 ± 0.98	8.34 ± 0.03
AHI (n/h)		43.00 ± 21.39	1.67 ± 2.60
TEST SET			
	All	OSAS Positive	OSAS Negative
Subjects (n)	83	49	34
Age (years)	58.09 ± 12.51	59.46 ± 12.20	56.15 ± 12.93
Males (%)	81.93	87.76	73.53
BMI (kg/m^2)	28.99 ± 5.11	30.07 ± 4.34	27.58 ± 5.75
Recording Time (h)	8.21 ± 0.69	8.25 ± 0.83	8.15 ± 0.43
AHI (n/h)		40.75 ± 20.47	1.68 ± 2.10

Table 2. Results provided on the test set by the selected neural network classifier and classic oximetry indices. Data are presented as mean ± standard deviation. Se: sensitivity; Sp: specificity; Ac: accuracy; AROC: area under receiving operating characteristic curve; MLP: multilayer perceptron; ODI2: oxygen desaturation over 2% index; ODI3: oxygen desaturation over 3% index; ODI4: oxygen desaturation over 4% index; CT90: cumulative time spent below 90% of saturation.

	Se (%)	Sp (%)	Ac (%)	AROC
MLP†	89.8	79.4	85.5	0.90 ± 0.04
MLP‡	90.8 ± 4.2	79.1 ± 6.9	86.0 ± 3.8	0.91 ± 0.03
ODI2	60.0	86.4	71.2	0.75 ± 0.07
ODI3	60.0	90.9	73.1	0.77 ± 0.07
ODI4	56.7	90.9	71.2	0.80 ± 0.06
CT90	40.0	95.5	63.5	0.78 ± 0.07

†: classification results were estimated by applying the hold-out method.
‡: classification results were estimated by applying the bootstrap method.

Figure legends

Fig. 1. Complete scheme of the neural network-based algorithm for OSAS diagnosis.

Fig. 2. Effect of varying the number of hidden nodes and the regularization parameter on the classification performance of our MLP networks.

Fig. 3. Hinton diagram of the selected MLP network. The size of the squares represent the magnitude of weights and biases. Colours white and black indicate a positive or negative value, respectively.

Fig. 4. ROC curves computed on the test set for the selected MLP network and classic oximetry indices.

Conflict of interest statement

There are no conflicts of interest that could inappropriately influence this research work.