Usefulness of Artificial Neural Networks in the Diagnosis and Treatment of Sleep Apnea-Hypopnea Syndrome

Daniel Álvarez, Ana Cerezo-Hernández,
Graciela López-Muñiz, Tania Álvaro-De Castro,
Tomás Ruiz-Albi, Roberto Hornero and
Félix del Campo

Additional information is available at the end of the chapter

Abstract

Sleep apnea-hypopnea syndrome (SAHS) is a chronic and highly prevalent disease considered a major health problem in industrialized countries. The gold standard diagnostic methodology is in-laboratory nocturnal polysomnography (PSG), which is complex, costly, and time consuming. In order to overcome these limitations, novel and simplified diagnostic alternatives are demanded. Sleep scientists carried out an exhaustive research during the last decades focused on the design of automated expert systems derived from artificial intelligence able to help sleep specialists in their daily practice. Among automated pattern recognition techniques, artificial neural networks (ANNs) have demonstrated to be efficient and accurate algorithms in order to implement computer-aided diagnosis systems aimed at assisting physicians in the management of SAHS. In this regard, several applications of ANNs have been developed, such as classification of patients suspected of suffering from SAHS, apnea-hypopnea index (AHI) prediction, detection and quantification of respiratory events, apneic events classification, automated sleep staging and arousal detection, alertness monitoring systems, and airflow pressure optimization in positive airway pressure (PAP) devices to fit patients' needs. In the present research, current applications of ANNs in the framework of SAHS management are thoroughly reviewed.

Keywords: sleep apnea-hypopnea syndrome, pattern recognition, automated biomedical signal processing, artificial neural networks, multilayer perceptron, feed-forward back-propagation, Bayes theory

1. Introduction

In their daily practice, physicians must frequently decide a definitive diagnosis or the most suitable treatment using several variables from multiple clinical data sources, which is a highly complex task. A huge amount of valuable healthcare-related information is currently available, from symptoms reported by the patient and details stored in their clinical history to biochemical data and outcomes from biomedical recordings or medical images. In this context, machine learning methods are essential to maximize the usefulness of medical data in order to expedite decisions and avoid misdiagnosis. In the last decades, the increasing development of computers and artificial intelligence has led to the use of decision support expert systems in the common clinical practice of several fields of medicine [1, 2]. The huge number of studies published in the context of biomedical engineering during the last years clearly shows this trend.

Bayesian theory was one of the first mathematical frameworks used to implement decision support systems. Regarding the classification of an item, according to the Bayes' decision rule, the predicted class must be the one that maximizes *a posteriori* probability in order to minimize the classification error. A major goal is to model the statistical characteristics of the problem under study, leading to expert systems able to assist physicians in decision-making processes. Among pattern recognition algorithms, conventional statistical classifiers, such as discriminant analysis [3] or logistic regression (LR) [4], and more recently artificial neural networks (ANNs) [5], have been widely applied. The widely known statistical classifiers assume that the class density function of input data is known *a priori*. Assumptions such as normal distribution, homoscedasticity, linearity, independency, or stationarity decrease the complexity of the classifier, minimize the classification error, and improve the performance. Nevertheless, these assumptions are not always consistent in real-world pattern classification problems, especially when working with limited datasets. Conversely, when using ANNs, no assumptions are made about the probability density functions of input features and the training data is used directly to optimize the decision rule [6]. Nevertheless, ANNs are characterized by a complex design stage. Both statistical and ANNs approaches have its advantages and limitations. However, the ability to model complex nonlinear problems, which are very common in biological systems, have made ANNs widely used in medical applications.

The first attempt to model information processing in biological systems by means of ANNs was carried out by McCulloch and Pitts in 1943 [7]. Since then, ANN-based algorithms have significantly evolved and their use in the field of medicine has increased considerably, particularly since the late 1990s. Some computer programs in the context of statistical medicine already include ANNs among their functionalities, which has contributed to increase their use in medical research. Nevertheless, "neural network" remains frequently a confusing term for many healthcare-related researchers. The implementation of an ANN has to be carried out by means of advanced software and some expertise is required to set up properly the user-dependent input parameters. However, once designed, they are reliable and easy to use tools, even by nontrained personnel. In addition, once optimized, the computational time is small, which is a major feature in order to speed up decision making.

Sleep research and particularly sleep-related breathing disorders (SBD) is a field in which the application of automated pattern recognition algorithms has increased exponentially during the last years due to the need for automating their complex diagnostic processes. Particularly challenging is the management of sleep apnea-hypopnea syndrome (SAHS). The gold standard technique for SAHS diagnosis is in-lab nocturnal polysomnography (PSG). During PSG, several neuromuscular and cardiorespiratory signals (up to 32 biomedical recordings) are monitored and stored for subsequent interpretation by trained personnel, which is a highly complex and time-consuming task [8]. In addition, accessibility to diagnosis and treatment is limited due to insufficient resources, both human (trained specialists) and technical (specialized sleep units), which have led to large waiting lists [9]. In this context, automated computer-aided diagnosis systems have emerged as very useful tools to deal with complex rules involving several biomedical recordings simultaneously, in order to expedite diagnosis and treatment [10–12]. Among all the machine learning-based tools, ANNs have been widely applied in the context of SAHS and merit a thorough analysis.

In order to analyze the usefulness of ANNs in the management of SAHS, an exhaustive review of the studies published during the past decade has been carried out. The review is structured as follows. First, the most relevant tasks regarding the ANNs learning process are outlined in Section 2. In this regard, some user-dependent decisions involving the ANN design and major issues concerning the training and testing processes are detailed. Second, in Section 3, the most relevant applications of ANNs are analyzed, including automated diagnosis, sleep staging, and treatment monitoring.

2. Artificial neural networks

ANNs are mathematical models inspired in the information processing capabilities of the nervous system designed to accomplish a predetermined task specified by the user [13, 14]. They were built to implement useful brain functions into a pattern recognition algorithm, such as parallel processing, distributed memory/storage, and environmental flexibility. ANNs are characterized by a fast and effective processing, learned from a preceding training process. During the learning or training stage, a wide set of known representative samples are used in order to model the statistical properties of the problem under study and accordingly compose the structure of the network. **Figure 1** illustrates a common network architecture of interconnected nodes arranged in layers simulating the brain's neuronal synapses.

The following advantages can be obtained when ANNs are applied for pattern recognition problems: (i) no prior assumptions about the data distribution are made as ANNs adjust themselves to the particular problem constrains during the learning process [15], (ii) ANNs are universal estimators able to match any function with arbitrary accuracy [16], and (iii) they are nonlinear algorithms able to model real-world complex relationships [15].

There are two major classes of ANNs: feedforward multilayered networks and radial basis function (RBF) networks. Both types of ANNs are capable of approximating any continuous functional mapping by means of several units (neurons) arranged in different layers [17]. The
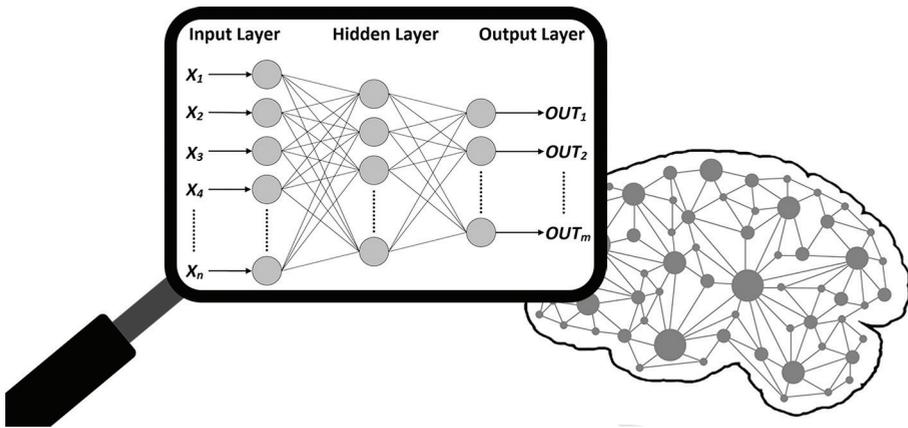
Figure 1. Common network architecture of interconnected nodes arranged in three layers (input, hidden, and output) simulating the brain's neuronal synapses.

main difference is the way hidden units are activated, i.e., how the input data is used to compute the output of each unit. In feedforward ANNs, there is a fixed (usually nonlinear) activation function, whereas in RBF ANNs, the activation of each unit depends on the radial distance (typically Euclidean) between the input vector and a prototype vector (center) [18].

The multilayer perceptron (MLP) is the most widely used feedforward ANN in computer-aided medical research. Indeed, feedforward networks, particularly MLP, are the most popular ANN in the framework of SAHS management [19–21]. A particular implementation of MLP networks involving Bayesian inference during the learning process (BY-MLP), which increase the generalization ability and allow for relevance analysis of input variables, has demonstrated to be useful in this context [22]. Similarly, probabilistic neural networks (PNN), which also integrates the Bayes' theory into the learning process, have been recently applied in the SAHS diagnosis problem [23]. In addition, RBF ANNs [24, 25], such as learning vector quantization (LVQ), which is a precursor of self-organizing maps using the Hebbian learning-based approach [26, 27]; fuzzy neural networks (FNN), which incorporate the fuzzy inference system (FIS) into the learning process [26, 28, 29]; self-organizing maps (SOM) and adaptive resonance theory (ART) models, which are likely the most common unsupervised ANNs [30, 31]; and recurrent neural networks (RNN), which allow for closed-loop connections between units (feedback) [32], have been also applied in the framework of automated SAHS management.

Next, an overview of the conventional multilayered network architecture is provided, as well as the most important issues regarding the design, training, and validation stages common to all approaches in the ANN-based framework. **Figure 2** shows a flow diagram summarizing these stages.

2.1. Network design: architecture of a neural network

The so-called neuron is the basic element within an ANN, which comprises its elementary mathematical functions [17]. ANNs are composed of multiple interconnected nodes arranged in different levels or layers leading to a massive parallel structure. The first level is called the
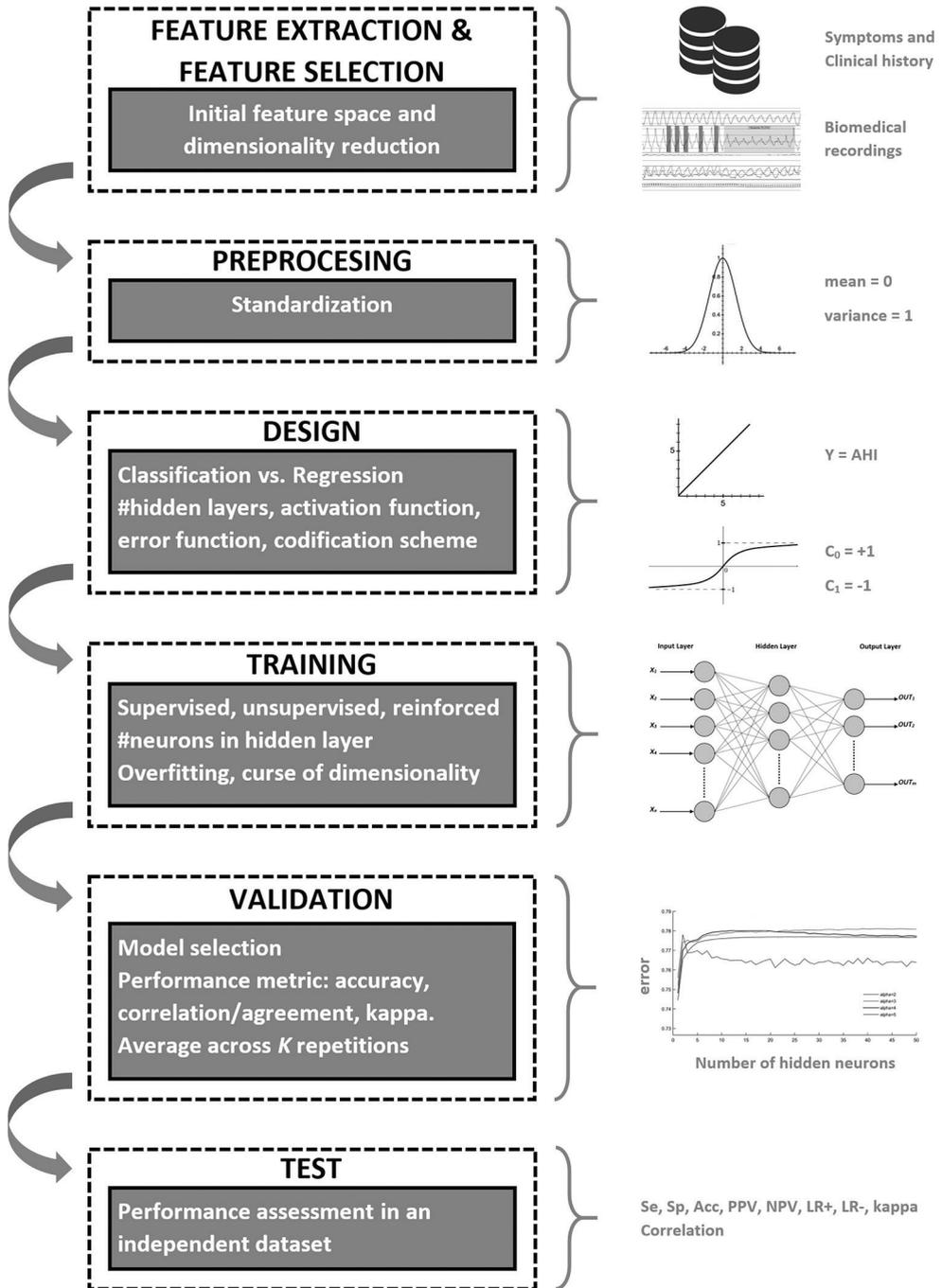
Figure 2. Flow diagram summarizing the training, validation, and test stages, as well as the most important issues involved in the design of an ANN.

input layer. Neurons in the input layer process directly the feature vectors or patterns that feed the ANN. Similarly, each output from every neuron in a layer feeds neurons composing the subsequent layer, leading to a distributed complex structure. The last level of the network, whose nodes provide the output of the ANN, is called the output layer. The remaining internal levels are called hidden layers. Both the number of hidden layers and the number of nodes are flexible and are determined during the learning process. The feedforward architecture is the most widely used, where each neuron in a layer is connected to every neuron on the next layer but neither connections between units in the same level nor closed-loops (feedback) are allowed. Therefore, data is always moving forward from one layer to the next, i.e., from the input to the output.

There is not a predetermined network architecture known to be *a priori* the best for any problem under study in terms of performance. The mathematical operation accomplished by each neuron is always the same. Therefore, the functionality of the ANN, i.e., the way in which a particular problem is addressed, is determined by the strength of the link between each pair of neurons. This strength is characterized by the coefficients of the ANN, the so-called weights, which are optimized during the training stage. Similar to the process of memory, weights represent the information stored in the network, whereas the optimization procedure represents the learning process or statistical inference [18].

As aforementioned, the structure of an ANN depends on the number of hidden layers, the number of neurons per layer, and the connectivity strength among them. Regarding the number of levels, it is common to construct ANNs with a single hidden layer because it has been demonstrated that this architecture is able to achieve universal approximation [33]. This is a user-dependent decision, whereas the number of neurons and the connectivity degree (weights) are both determined automatically during the learning process. Regarding the number of nodes in the hidden layer, it is commonly optimized by means of a hold-out or cross-validation approach using the data in a training dataset. In this regard, it is supposed that the complex the problem, the higher the number of neurons. Notwithstanding, even a small network with a reduced number of nodes can model complex problems and reach high prediction ability. In addition, the following design issues must be addressed before the learning process [17]: the output coding scheme, the error function used in the network training, and the activation function of neurons in the hidden and output layers. The hyperbolic tangent function is a common activation function for neurons in the hidden layer since it has been demonstrated that it provides fast convergence of training algorithms [13, 17]. **Figure 3** shows a common schema of a single neuron (perceptron) with a sigmoid activation function. Regarding the learning process, the scale conjugate gradient (SCG) is a common method for updating the adjustable parameters of the ANN (weights and biases) during the training stage.

2.1.1. Classification and regression approaches

According to the mathematical nature of the output, ANNs can be applied to address two main kinds of problems: classification and regression. Regarding the classification approach, the goal of the ANN is to estimate the class membership for an input feature pattern among a
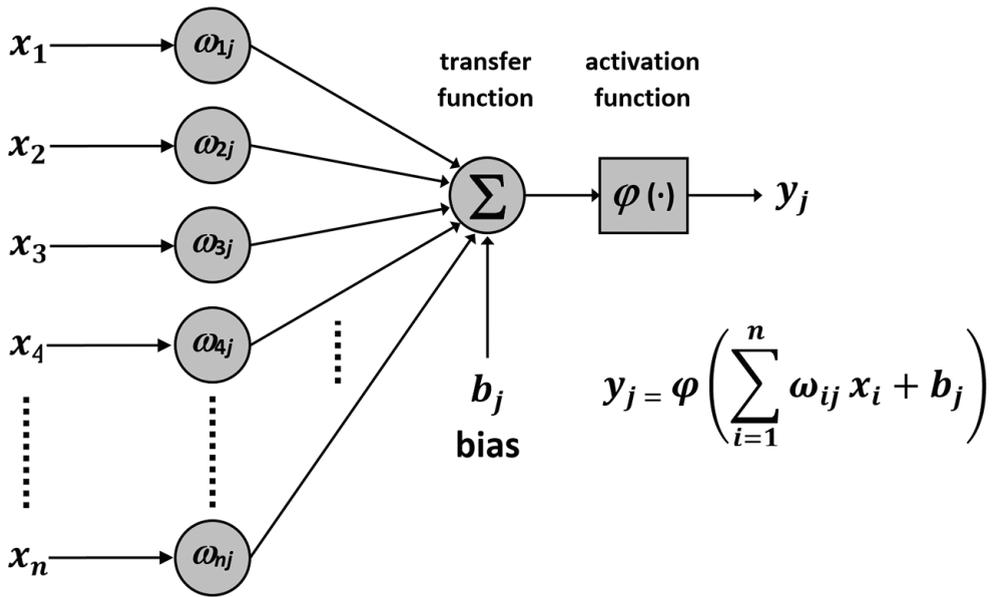
Figure 3. Scheme of a perceptron. A nonlinear activation function $\phi(\cdot)$ is applied to the weighted sum of the input features (x_n) and the bias term (b_j) in order to compute the output (y_k).

set of predefined discrete categories. Conversely, in a regression task, the goal of the ANN is to estimate a continuous variable.

In the context of binary classification problems, an output layer with just a single neuron is needed. Regarding, for instance, a 2-class SAHS diagnosis problem, all input patterns are assigned to one of two mutually exclusive classes: SAHS positive (class C_0 or positive class) or SAHS negative (class C_1 or negative class). A possible target coding scheme would be the following: $t = 0$ for the positive class and $t = 1$ for the negative class. This architecture can be used also in regression problems, where the variable to be approximated is unidimensional and continuous. In the context of SAHS diagnosis, the goal of a regression ANN could be to estimate the apnea-hypopnea index (AHI).

Due to a highly flexible architecture, most of the ANNs can be used to model both classification and regression problems by just modifying certain design characteristics [17]. The main difference between classification and regression ANNs is linked with the nature of the function to be approximated. The output of an ANN is provided in terms of probability in a classification task while it is an estimate of a continuous variable in a regression context. Accordingly, optimization procedures differ from one approach to another. Regarding a binary classification approach, the activation function of the output neuron could be a nonlinear function with output values ranging [0, 1], e.g., sigmoid functions such as the logistic or the hyperbolic tangent. In this regard, the network output can be interpreted as the probability that the input feature

pattern belongs to one class or another according to the Bayes' theorem. Conversely, addressing a regression task, the network output values must be continuous and nonnegative. Therefore, a linear activation function ranging $[0, \infty)$ would be suitable.

Regarding the error function governing the learning process, the cross-entropy error function is widely used in the context of binary classification, whereas the sum-of-squares error function is commonly used for regression purposes [17].

2.1.2. Standardization of input patterns

Normalization of input feature values is an important task in nonlinear pattern recognition methods [34]. Bounded similar input magnitudes are needed to accomplish suitable weight initialization. Input patterns are composed of features parameterizing different properties of the problem under study, e.g., the influence of recurrent apnea events typical of SAHS on cardiorespiratory signals. Usually, several features of different nature are involved in order to obtain as much information as possible, e.g., sociodemographic, anthropometric, clinical, and/or variables from automated feature extraction algorithms. Therefore, their values may differ significantly and thus they must be normalized. In this regard, simple linear rescaling can be used to standardize (zero mean and unit variance) the magnitudes of each input feature by subtracting its mean and dividing by its standard deviation.

2.2. The training process: learning the problem under study

Training is the most important stage when working with ANNs. The aim of the training process is to adapt the ANN to the problem under study by computing some adjustable parameters. The training or learning process can be (i) supervised, in which the learning process is guided by a static mapping between input patterns and known targets; (ii) reinforced, in which a performance function assesses the accuracy of the current output instead of knowing the actual target values; and (iii) unsupervised, in which ANNs adapt themselves to input patterns with no kind of feedback [35]. In the context of medical decision support systems, the supervised approach is the most widely used. When using supervised learning, it is essential to know the target or actual output value for a wide set of input patterns. The dataset of examples used during the learning stage is referred to as the training set. According to this training input-output pairs, the network weights are tuned to fit the input to its corresponding target. It is important that the training set would be large enough to represent fairly the problem under study.

The backpropagation learning is the most commonly used methodology for updating weights in feedforward ANNs due to its computational efficiency [17]. Using this approach, all weights are updated every time an input pattern is fed from the training dataset in order to minimize an error function. First, the network weights are initialized randomly. During a supervised learning process, the training samples (input-target pairs) are fed into the network and the error function is computed, i.e., the difference between the estimated output value and the desired target according to a predefined suitable function. Then, the values of the network weights are modified in order to minimize the error. This procedure is repeated throughout several iterations, which are set by the user. Once the training process is finished, all network weights already have a fix value, i.e., there is a single optimized ANN able to carry out the task for which it was designed.

2.2.1. Generalization ability and the problem of overfitting

Once optimized, an ANN is able to process new input patterns independent of the training dataset. In this regard, it is noteworthy that the goal of the training stage must be to build a general statistical model of the problem under study rather than to learn data samples from a particular training set. This is an essential characteristic common to all pattern recognition techniques and it is required to achieve good generalization ability. Generalization accounts for the ability to make good predictions for new unknown inputs [17].

In addition to the user's capability to accomplish appropriate design and optimization procedures, the performance or generalization ability of an ANN is influenced by three main factors [13, 36]: (i) the size and completeness of the training dataset, i.e., whether the learning samples account for all the variability of the environment or problem of interest; (ii) the number of adjustable parameters in the model; and (iii) the complexity of the problem under study. The nature of the problem or model complexity is linked with the number of adjustable parameters in the ANN (network weights) and it cannot be controlled. Theoretically, the harder the problem, the more complex the ANN. In this regard, it is important to achieve a compromise between the generalization ability and complexity. An ANN with a small number of parameters, i.e., low flexibility, may lead to an underfitted model, insufficient to reach high generalization. On the contrary, an ANN with a large number of weights may lead to an overfitted model that matches a particular training dataset, resulting in poor generalization. Underfitting can be avoided by increasing the flexibility, whereas overfitting requires the training set to grow accordingly to the network complexity [13].

In the same way, the optimization of an ANN is closely related to the bias-variance trade-off. A too simple or inflexible model will have a large bias and may lead to underfitting. Conversely, models with a high variance provide high flexibility but could adapt to the noise present in the training set, leading to overfitting. Bias and variance are both complementary characteristics and thus the best generalization is obtained when a compromise between the conflicting requirements of small bias and small variance is achieved [15, 17].

A way to reduce both bias and variance simultaneously is to increase the number of training samples. As a result, model complexity increases, which minimizes the bias. At the same time, constrains imposed by the training data will be more rigorous, thereby also reducing variance. As mentioned earlier, to achieve this goal the size of the training set should increase in accordance with model complexity [17]. Nevertheless, this requirement cannot always be achieved in real-world applications because the size of the training set is usually fixed and limited. Therefore, finding the optimum model complexity is a major issue. In order to deal with this optimization problem, a new trade-off arises: simpler models are preferred but smoothing mapping is needed to prevent from poor generalization [13, 17]. In this regard, regularization techniques allow the ANN to control the effective complexity of the model by reducing the number of adjustable parameters during the training set. Weight decay and early stopping are common approaches of regularization. Weight decay is probably the most widely used, consisting on adding a penalty term to the error function in order to penalize complex mappings.

An additional issue regarding the training sample size is called the course of dimensionality [17]. This term refers to the relationship between the size of the training set and the dimension of the feature space, i.e., the number of variables in the input feature vector. The course of dimensionality states that the number of training samples needed to characterize the underlying problem grows exponentially as the number of input features increases. Therefore, the size of the training dataset must also increase according to the input space dimension in order to enhance generalization ability and avoid overfitting [18].

As previously stated, the size of the training set in real-world applications is fixed and usually limited, especially in the field of medicine. In this regard, dimensionality reduction techniques contribute to address the problem of overfitting due to the curse of dimensionality. An ANN fed with fewer input features needs to optimize fewer parameters (weights) and these are more likely to be properly characterized by a limited training dataset. The aim of dimensionality reduction algorithms is to compose a reduced subset of the most significant features governing a model. To achieve this goal, a fitness metric (relevancy, redundancy, completeness, or accuracy, among others) is used to obtain the optimum feature subset. There are several feature selection methodologies but principal component analysis and stepwise feature selection are likely the most widely used in medical applications.

2.3. Validation and test processes: model selection and performance assessment

In order to estimate the actual prediction ability of an ANN, the learning, model selection, and performance assessment stages must be carried out using independent datasets, i.e., the so-called training, validation, and test datasets. The goal of model selection is to obtain the optimum network configuration by comparing the performance of several ANNs with different values of the design parameters, i.e., number of neurons in the hidden layer and usually the regularization parameter. The hold-out method is commonly used for this purpose because it avoids a biased estimation of the results [36]. In the hold-out method, the initial population/dataset is split into three independent groups for training, validation, and testing purposes. The network weights are adjusted in the training set for different configurations of the adjustable parameters specified by the researcher, i.e., multiple ANNs are really trained, whereas the performance of each individual ANN is computed in the validation set to determine the optimum ANN for the problem under study. Since there is a random initialization of weights, the training process is frequently repeated several times to avoid a potential bias linked with this arbitrary decision. Thus, the performance metric for model selection from the validation set is averaged across all the repetitions. Nevertheless, this procedure can lead also to some overfitting so the selected optimum ANN has to be further assessed in an independent test set composed of unseen data samples [17].

It is worth to notice that, unfortunately, several studies from the literature do not implement a suitable validation of their proposed methodology, providing biased overoptimistic results [37]. On the other hand, sometimes the initial dataset is not large enough to properly derive the three independent subpopulations. In such cases, cross-validation techniques allow for training and validating the models in the same training set without biasing the selection of the optimum model. Bootstrap, leave-one-out, and k-fold cross-validation are common algorithms to deal with small populations under study.

3. Clinical applications of NNs in the context of sleep apnea-hypopnea syndrome

ANNs have been applied to model problems in several fields, such as industrial processes optimization, economic and financial modeling, chemistry, physics, biology, or medicine, among others [38–42]. In the framework of SAHS management, automated expert systems based on ANNs have been mainly applied to classify patients suspected of suffering from SAHS (binary classification: no SAHS vs. SAHS), to categorize the severity of the disease (multiclass classification: no SAHS, mild, moderate, and severe), to estimate the AHI (regression of a continuous variable), to detect and quantify respiratory events (normal breathing vs. apneic), and to categorize apneic events (central, obstructive, and mixed). ANNs have been also used to implement automated sleep staging and arousal detection, which are very useful functionalities incorporated in current commercial software applications for sleep analysis. In addition, ANNs play an important role in alertness monitoring systems and they are already integrated in positive airway pressure (PAP)-based treatment devices to fit user's airflow needs, which are major issues for patients suffering from SBDs.

Most research in the field of SAHS focus on binary classification in order to determine the presence or absence of the disease. Similarly, some studies also applied ANNs for multiclass classification in order to characterize SAHS severity according to predefined discrete categories. Conversely, despite of its higher information about the severity of the disease, only a few studies have been carried out to estimate the AHI using a regression approach (continuous function).

Regarding the nature of the input data, ANNs aimed at assisting in SAHS diagnosis first used anthropometric and clinical features to compose input patterns [19, 43]. However, the increasing research in the context of biomedical signal processing allows physicians to derive essential information directly from signals monitored during the PSG [44]. In this regard, blood oxygen saturation (SpO_2) from oximetry and heart rate variability (HRV) from electrocardiogram (ECG) are the most widely used. In addition, airflow from both thermistor and nasal pressure, abdominal and chest movements, snoring sounds, and EEG have been also studied. Alternatively, in order to avoid sleep studies, automated signal processing of speech recordings and even image analysis for facial recognition have been also assessed as an alternative to PSG-derived signals to assist in the detection of SAHS.

The main goal of computer-aided tools for SAHS management is to simplify and speed up the diagnostic methodology, in order to alleviate large waiting lists and increase accessibility of patients to diagnostic resources. Current research focuses on analyzing a reduced set of biomedical recordings, which are preferably obtained at patient's home using existing commercial portable devices. Therefore, powerful tools are needed to obtain as much information as possible from this reduced subset of signals. In this regard, ANNs allow researchers to manage several features derived from the signals under study and thus they are suitable and reliable tools to help physicians in the diagnosis of SAHS. In order to obtain complementary information, different automated signal processing methods have been applied, such as common statistics (mean, median, variance, skewness, kurtosis), time domain analyses (detection and quantification of

respiratory events), frequency domain analyses (Fourier analysis, time-frequency maps, wavelet transform, bispectrum), and/or nonlinear methods (entropy measures, Poincaré plots, complexity measures), among others, both individually or jointly.

3.1. SAHS diagnosis by means of ANNs

ANNs were first used in the context of SAHS detection in the late 1990s, when Kirby et al. [43] and El-Solh et al. [19] carried out retrospective analyses aimed at designing ANNs based on clinical and anthropometric variables from patients showing clinical suspicion of SAHS. **Table 1** summarizes the main characteristics of significant studies carried out during the last decade focused on applications of ANNs aimed at assisting in SAHS diagnosis. In the study by Kirby et al. [43], 23 clinical variables fed a generalized regression neural network (GRNN), which is a kind of RBF network, for binary classification (SAHS vs. no SAHS). The authors reported 98.9% sensitivity, 80.0% specificity, and 91.3% accuracy (86.8–95.8, CI 95%). Similarly, El-Solh et al. [19] used clinical and anthropometric variables in order to estimate the AHI by means of a MLP ANN. Using cutoffs of 10, 15, and 20 events per hour (e/h) for a positive diagnosis of SAHS, the sensitivity-specificity pairs were 94.9–64.7%, 95.3–60.0%, and 95.5–73.4%, respectively. Both studies achieved significantly high sensitivity but poor to moderate specificity, which is a common trend of pattern recognition techniques in the context of SAHS.

Recent studies have built updated predictive models based on anthropometric and clinical data, since characteristics of patients referred nowadays to sleep units have changed compared to those of patients in the last decade. In this regard, Su et al. [45] proposed the multiclass Mahalanobis-Taguchi system (MMTS) and used both anthropometric information and questionnaire data in order to classify patients into normal subjects or mild, moderate, or severe SAHS patients. Additionally, LR, conventional feed-forward backpropagation FFBB and LVQ ANNs, support vector machines (SVM), C4.5 decision tree (DT), and rough set (RS) were also applied for comparison purposes. The proposed MMTS significantly outperformed the competing classifiers, reaching an average accuracy of 84.38% (normal: 87.50%; mild: 66.67%; moderate: 100%; severe: 83.33%). Particularly, FFBB and LVQ ANNs reached 34.04% (normal: 25.00%; mild: 33.33%; moderate: 11.11%; severe: 66.70%) and 47.22% (normal: 50.00%; mild: 16.67%; moderate: 22.22%; severe: 100%) overall accuracy, respectively. Similarly, in a recent study carried out by Wang et al. [27] several automated classifiers fed with anthropometric and questionnaire-based variables were also assessed to predict SAHS. The authors propose a novel classifier based on fuzzy decision trees (FDT) to detect SAHS. In addition, LR, ANNs (backpropagation and LVQ), a SVM, and a conventional DT were used as benchmarks for comparison purposes. The proposed FDT achieved the highest performance (81.82% accuracy, 0.554 *kappa*, and 0.673 geometric mean). However, a synthetic oversampling approach (SMOTE) was used to deal with the common imbalance between SAHS positive and SAHS negative classes, which was not used in the remaining benchmark methods. Without SMOTE, FDTs slightly outperformed the backpropagation ANN (48.22% vs. 47.53% accuracy, 0.186 vs. 0.175 *kappa*, and 0.300 vs. 0.288 geometric mean), whereas the highest precision was achieved by the conventional LR approach (49.57% accuracy, 0.207 *kappa*, and 0.320 geometric mean). Karamanli et al. recently assessed a MLP ANN trained to classify healthy and SAHS patients using sex, age, BMI, and snoring status as input variables, reporting 86.6% accuracy [21]. Nevertheless, it is important to highlight that input features derived automatically from

Author (year) [Ref.]	ANN model	Purpose	Target function/ class(es)	Input variables	Performance metrics
Kirby et al. (1999) [43]	GRNN	Classification (binary)	No SAHS vs. SAHS (AHI ≥10 e/h)	Clinical	98.9% Se 80.0% Sp 91.3% Acc
El-Solh et al. (1999) [19]	MLP	Regression	AHI estimation	Clinical and anthropometric	CC = 0.852 cutoff 10 e/h 94.9% Se - 64.7% Sp cutoff 15 e/h 95.3% Se - 60.0% Sp cutoff 20 e/h 95.5% Se - 73.4% Sp
Su et al. (2012) [45]	MMTS LR FFBB LVQ SVM DT RS	Classification (4-class)	Normal/mild/ moderate/severe	Anthropometric and questionnaire data	84.38% average Acc 55.33% average Acc 34.04% average Acc 47.22% average Acc 53.82% average Acc 63.54% average Acc 13.20% average Acc
Wang et al. (2016) [27]	FFBB LVQ	Classification (4-class)	No SAHS/mild/ moderate/severe	Anthropometric and questionnaire data	47.5% Acc, 0.145 k, 0.288 g-mean 43.4% Acc, 0.181 k, 0.280 g-mean
Karamanli et al. (2016) [21]	MLP	Classification (binary)	No SAHS vs. SAHS (AHI ≥10 e/h)	Sex, age, BMI, snoring status	86.6% Acc
Polat et al. (2008) [29]	FFBB ANFIS	Classification (binary)	No SAHS vs. SAHS (AHI ≥ 5 e/h)	In-lab PSG-derived	100% Se, 93.5% Sp, 95.1% Acc, 0.96 AUC
Ghandeharioun et al. (2015) [30]	SOM	Classification (4-class)	No SAHS/mild/ moderate/severe	In-lab PSG-derived and anthropometric	94.2% Se, 97.8% Sp, 96.5% Acc
Marcos et al. (2008) [20]	MLP	Classification (binary)	No SAHS vs. SAHS (AHI ≥10 e/h)	Nonlinear features from SpO_2	89.8% Se, 79.4% Sp, 85.5% Acc
Marcos et al. (2008) [24]	RBF-KM RBF-FCM RBF-OLS	Classification (binary)	No SAHS vs. SAHS (AHI ≥10 e/h)	Nonlinear features from SpO_2	89.4% Se, 81.4% Sp, 86.1% Ac 86.6% Se, 81.9% Sp, 84.7% Acc 89.8% Se, 79.4% Sp, 85.5% Acc
Almazaydeh et al. (2012) [46]	MLP	Classification (binary)	Healthy vs. SAHS (AHI ≥5 e/h) Physionet	ODI3, delta index, CTM from SpO_2	87.5% Se, 100% Sp, 93.3% Acc
Marcos et al. (2010) [22]	MLP BY-MLP	Classification (binary)	No SAHS vs. SAHS (AHI ≥10 e/h)	Statistical, spectral, and nonlinear features from SpO_2	86.4% Se, 62.8% Sp, 76.8% Acc 87.8% Se, 82.4% Sp, 85.6% Acc
Morillo et al. (2012) [47]	BY-MLP	Classification (binary)	No SAHS $vs.$ SAHS (AHI ≥10 e/h)	Time, stochastic, spectral, and nonlinear features from SpO_2	92.4% Se, 95.9% Sp, 93.9% Acc

Author (year) [Ref.]	ANN model	Purpose	Target function/ class(es)	Input variables	Performance metrics
Huang et al. (2015) [26]	FFBB	Classification (binary)	No SAHS vs. SAHS (AHI ≥5 e/h)	ODI4 from SpO$_2$	88.0% Se, 93.3% Sp, 90.7% Acc
	LVQ				80.7% Se, 79.3% Sp, 80.0% Acc
	ANFIS				90.7% Se, 86.0% Sp, 88.3% Acc
Khandoker et al. (2009) [23]	SVM	Classification (binary)	Healthy vs. SAHS (AHI ≥5 e/h) Physionet	Wavelet decomposition of HRV and EDR from ECG	100% Se, 100% Sp, 100% Acc
	LDA				90% Se, 100% Sp, 93% Acc
	KNN				80% Se, 90% Sp, 83% Acc
	PNN				80% Se, 50% Sp, 70% Acc
Khandoker et al. (2008) [48]	FFBB	Classification (binary)	Apneic vs. Normal Hypopnea vs. Apnea Obstructive vs. Central	ECG	87.6% Se, 95.5% Sp, 95.1% Acc
					86.1% Se, 78.7% Sp, 83.4% Acc
					93.7% Se, 99.2% Sp, 98.9% Acc
Acharya et al. (2011) [49]	FFBB	Classification (3-class)	Normal/apnea/ hypopnea	Nonlinear measures from ECG	95.0% Se, 100% Sp, 99.1% Acc (normal)
					88.0% Se, 90.0% Sp, 96.5% Acc (apnea)
					80.0% Se, 89.5% Sp 87.8% Acc (hypopnea)
Lweesky et al. (2011) [50]	FFBB	Classification (binary)	Normal breathing vs. apnea epochs	P-wave features from ECG	90.0% Se, 94.2% Sp, 92.0% Acc
Mendez et al. (2009) [51]	FFBB	Classification (binary)	Normal breathing vs. apnea (AHI ≥5 e/h)	Time and spectral features from RRi and QRS area time series	89.0% Se, 86.0% Sp, 88.0% Acc (m-by-m) 100% Acc (record)
Nguyen et al. (2014) [52]	ANN	Classification (binary)	Normal sleep vs. sleep apnea epochs	HRV complexity by means of RQA	85.6% Se, 79.1% Sp, 83.2% Acc
	SVM				93.7% Se, 65.9% Sp, 84.1% Acc
	Ensemble				86.4% Se, 83.5% Sp, 85.3% Acc
Fiz et al. (2010) [53]	MLP	Classification (binary)	No SAHS vs. SAHS AHI ≥5 e/h AHI ≥15 e/h	Time and spectral features from snoring recordings	87.0% Se, 71.0% Sp 80.0% Se, 90.0% Sp
Nguyen and Won (2015) [54]	f-MLP MLP	Classification (binary)	Normal breathing vs. snoring	Spectral content snoring recordings	96.0% overall Acc 82.0% overall Acc
Tagluk et al. (2011) [55]	MLP	Classification (binary)	Normal vs. SAHS EEG epochs	Bispectral analysis of EEG	94.1% Se, 98.2% Sp, 96.2% Acc

Author (year) [Ref.]	ANN model	Purpose	Target function/ class(es)	Input variables	Performance metrics
Liu et al. (2008) [31]	ART2	Classification (binary)	Healthy vs. SAHS subjects (AHI ≥5 e/h)	EEG energy in theta (Fourier transform) and pupil size	91.0% Acc
Lin et al. (2006) [28]	FFBB	Classification (binary)	No SAHS vs. SAHS epochs	EEG power in delta, theta, alpha, and beta using DWT	69.64% Se, 44.44% Sp
Akşahin et al. (2012) [56]	FFBB RBF DTD	Classification (3-class)	Obstructive/central/ healthy patients	Coherence and mutual information of EEG	0.1450 MRAE error 0.3692 MRAE error 0.2282 MRAE error
Fontela et al. (2005) [57]	BY-MLP	Classification (3-class)	Obstructive/central/ mixed	Wavelet decomposition of thoracic effort	83.78% Acc (overall) 80.90% Acc (obstr.) 89.95% Acc (centr.) 80.48% Acc (mixed)
Tagluk et al. (2010) [58]	FFBB	Classification (3-class)	Obstructive/central/ mixed	Wavelet decomposition of abdominal effort	78.50% Acc (overall) 73.42% Acc (obstr.) 94.23% Acc (centr.) 66.16% Acc (mixed)
Berdiñas et al. (2012) [59]	Ensemble ANNs	Classification (3-class)	Obstructive/central/ mixed	Wavelet decomposition of thoracic effort	90.27% Acc (overall) 94.62% Acc (obstr.) 95.47% Acc (centr.) 90.45% Acc (mixed)
Weinreich et al. (2008) [60]	FFBB	Classification (4-class)	OA/OH/CSR/normal breathing	Spectral entropy of airflow	91.5% Acc (overall) 90.2% Se, 90.9% Sp (OA vs. CSR) 91.3% Se, 94.6% Sp (OH vs. normal)
Várady et al. (2002) [61]	FFBB	Classification (3-class)	Normal/apnea/ hypopnea	IRA and IRI from airflow and RIP	93.0% Acc (overall) 98.4% Se, 94.0% Sp (normal) 78.7% Se, 91.0% Sp (hypopnea) 97.0% Se, 88.7% Sp (apnea)
Belal et al. (2011) [62]	MLP	Classification (binary)	Non-apneic vs. apneic event	Correlation and PCA of HR, RR, and SpO_2	81.8% Se, 75.8% Sp, 76.8% Acc
Marcos et al. (2012) [63]	MLP	Regression	AHI estimation	Spectral and nonlinear features from SpO_2	ICC = 0.91 cutoff 5 e/h 91.8% Se–58.8% Sp cutoff 10 e/h 89.6% Se–81.3% Sp cutoff 15 e/h 94.9% Se–90.9% Sp

Author (year) [Ref.]	ANN model	Purpose	Target function/ class(es)	Input variables	Performance metrics
Gutiérrez-Tobal et al. (2013) [25]	MLP RBF	Regression	AHI	Statistical, spectral, nonlinear features from airflow	ICC = 0.849 ± 0.002 cutoff 10 e/h 92.5% Se, 89.5% Sp, 91.5% Acc ICC = 0.748 ± 0.037 cutoff 10 e/h 92.5% Se, 57.9% Sp, 81.4% Acc
de Silva et al. (2011) [64]	FFBB	Regression	AHI	Pitch, formant, and structure-based features from snoring sounds and the neck circumference	Cutoff 15 e/h 91 ± 6% Se, 89 ± 5% Sp Cutoff 30 e/h 86 ± 9% Se, 88 ± 5% Sp
de Silva et al. (2012) [65]	FFBB	Regression	AHI	Pitch, formant, and structure-based features from snoring sounds and the neck circumference	Female, AHI ≥ 15 e/h 91 ± 10% Se, 88 ± 5% Sp Male, AHI ≥ 15 e/h 91 ± 6% Se, 89 ± 5% Sp Comb., AHI ≥ 15 e/h 84 ± 10% Se, 83 ± 13% Sp
Emoto et al. (2012) [66]	MLP	Regression	Breathing sound signal	Preceding samples of the breathing signal	89.2% average Se 87.4% average Sp

Notes: Se: sensitivity; Sp: specificity; Acc: accuracy; e/h: events per hour; CC: correlation coefficient; k: kappa coefficient; g-mean: geometric-mean; ICC: intra-class correlation coefficient; GRNN: generalized regression neural network; SAHS: sleep apnea-hypopnea syndrome; AHI: apnea-hypopnea index; MLP: multilayer perceptron; MMTS: multiclass Mahalanobis-Taguchi system; LR: logistic regression; FFBB: feed-forward back-propagation; LVQ: learning vector quantization; SVM: support vector machine; DT: decision tree; RS: rough set; ANFIS: adaptive network-based fuzzy inference system; SOM: self-organizing maps; BMI; body mass index; RBF: radial basis function; KM: k-means; FCM; fuzzy c-means; OLS: orthogonal least squares; SpO_2: blood oxygen saturation from nocturnal oximetry; ODI3: oxygen desaturation index of 3%; CTM: central tendency measure (nonlinear); BY-MLP: Bayesian training MLP neural network; PNN: probabilistic neural network; ODI4: oxygen desaturation index of 4%; HRV: heart rate variability; EDR: ECG-derived respiration; RRi: R-to-R interval time series; QRS: QRS complex from the ECG; k-NN: k nearest neighbors; RQA: recurrence quantification analysis; f-MLP: correlational filter MLP; ART2: modified adaptive resonance theory ANN; DWT: discrete wavelet transform; OA: obstructive apnea; OH: obstructive hypopnea; CSR: Cheyne-Stokes respiration; DTD: distributed time-delay neural network; MRAE: mean relative absolute error; IRA: instantaneous respiration amplitude; IRI: instantaneous respiration interval; RIP: respiratory inductance plethysmography; HR: heart rate; RR: respiratory rate.

Table 1. Performance and the most relevant characteristics of the studies using ANNs in the context of SAHS classification, event detection, and AHI regression.

cardiorespiratory and/or neuromuscular signals have been used predominantly, while anthropometric and clinical variables have been used marginally.

In the study by Polat et al. [29], different expert systems were assessed to classify patients with suspicion of SAHS using clinical features derived from in-lab polysomnography, including the arousal index and the AHI. A FFBB ANN reached 100% sensitivity, 93.55%

specificity, 95.12% accuracy, and 0.96 AUC, slightly lower and more unbalanced than a DT-based classifier (91.67% sensitivity, 96.55% specificity, 95.12% accuracy, and 0.97 AUC). This work assessed the usefulness of different expert systems in the context of SAHS, although using input variables computed from the whole PSG study limits its ability as screening test for the disease. Similarly, Ghandeharioun et al. [30] trained a 4-class SOM to classify patients suspected of suffering from SAHS into healthy, mild, moderate, and severe categories using PSG-derived and anthropometric variables. The proposed algorithm reached 94.2% sensitivity, 97.8% specificity, and 96.5% accuracy, although neither validation nor test stages were described.

Regarding SAHS diagnosis by means of ANNs, the SpO_2 signal from nocturnal oximetry is probably the most widely used biomedical data source. In the study by Marcos et al. [20], approximate entropy (ApEn), central tendency measure (CTM), and Lempel-Ziv complexity were applied to the SpO_2 nocturnal profile to estimate irregularity, variability, and complexity, respectively. These nonlinear measures composed the input feature patterns to feed a MLP ANN for SAHS binary classification. A sensitivity of 89.8%, specificity of 79.4%, and accuracy of 85.5% were obtained in an independent test set, significantly improving the diagnostic performance of conventional oximetric indices. The same authors reached similar diagnostic performance using a RBF ANN in the same context [24]: average accuracies of $86.1 \pm 1.1\%$ ($89.4 \pm 1.6\%$ sensitivity, $81.4 \pm 1.7\%$ specificity), $84.7 \pm 1.2\%$ ($86.6 \pm 2.8\%$ sensitivity, $81.9 \pm 2.0\%$ specificity), and $85.5 \pm 0.0\%$ ($89.8 \pm 0.0\%$ sensitivity, $79.4 \pm 0.0\%$ specificity) were achieved using k-means, fuzzy c-means, and orthogonal least squares kernels, respectively. An MLP ANN was also assessed in the study by Almazaydeh et al. [46] to perform binary classification. The ANN was fed with the conventional oxygen desaturation index of 3% (ODI3), the delta index, and the CTM from overnight oximetry recordings, reaching 87.5% sensitivity, 100% specificity, and 93.3% accuracy in a test set from the publicly available *PhysioNet* dataset.

Bayesian training has been applied to deal with overfitting of ANNs. In addition, Bayesian inference also allows the user to measure quantitatively the influence of each input feature in the output of the model. The effectiveness of this approach was assessed in the study by Marcos et al. [22]. A sensitivity of 87.76%, specificity of 82.39%, and accuracy of 85.58% were reached, significantly improving the performance achieved using the conventional maximum likelihood criterion (86.42% sensitivity, 62.83% specificity, and 76.81% accuracy). Similarly, Sánchez-Morillo et al. [47] applied a feedforward probabilistic ANN to classify patients into SAHS negative or SAHS positive using time, stochastic, spectral, and nonlinear features from nocturnal SpO_2 recordings. A sensitivity of 92.42%, specificity of 95.92%, and accuracy of 93.91% were reached in a single training set using leave-one-out cross-validation. In a recent study by Huang et al. [26], the automated analysis of the oxygen desaturation index of 4% (ODI4) from oximetry by means of a DT was proposed as an abbreviated method for SAHS screening. In this work, the authors assessed several pattern recognition techniques for automated diagnosis, including some ANNs, such as conventional backpropagation, (LVQ), and adaptive network-based fuzzy inference system (ANFIS). The proposed DT reached 98.67% sensitivity, 90.67 specificity, and 94.67% accuracy, outperforming backpropagation (88.00% sensitivity, 93.33% specificity, 90.67% accuracy), ANFIS (90.67% sensitivity, 86.00% specificity,

88.33% accuracy), and LVQ (80.67% sensitivity, 79.33% specificity, 80.00% accuracy) ANNs. In this study, conventional LR and k-nearest neighbors (k-NN) combined with genetic algorithms (GAs) and particle swarm optimization (PSO) also outperformed ANNs.

ECG recordings have been also widely used to assist in SAHS diagnosis. In the study by Khandoker et al. [23], the spectral content of HRV and ECG-derived respiration (EDR) time series from single-lead ECG recordings were analyzed by means of the wavelet transform. The authors proposed a binary SVM for classification (healthy vs. SAHS) and compared its performance with LDA, k-NN, and PNN. The proposed SVM classifier reached 100% accuracy in the test set, whereas the PNN showed poor classification performance (80% sensitivity, 50% specificity, and 70% accuracy) probably due to a suboptimal setting of the spread parameter (σ) of the Gaussian function. In a previous study by Khandoker et al. [48], the authors analyzed ECG short-term epochs from nocturnal PSG by means of wavelet decomposition to classify segments into normal breathing, obstructive apnea, and central apnea using a feedforward ANN. The authors reported accuracies of 95.10% in the classification of apneic and normal breathing epochs, 83.40% in the detection of hypopneas, and 98.96% in the classification of obstructive and central apneas. Similarly, Acharya et al. [49] implemented a FFBB ANN using nonlinear measures from the ECG (ApEn, fractal dimension, correlation dimension, largest Lyapunov exponent, and Hurst exponent) to detect apneas, hypopneas, and normal breathing segments. The proposed ANN reached 99.1% accuracy (95.0% sensitivity, 100% specificity), 96.5% accuracy (88.0% sensitivity, 90.0% specificity), and 87.8% accuracy (80.0% sensitivity, 89.5% specificity) in the classification of normal breathing, apneas, and hypopneas, respectively. Lweesky et al. [50] focused on the characterization of the P-wave of the ECG in order to feed an ANN aimed at discerning between apnea and normal breathing. The authors reported 90.0% sensitivity, 94.2% specificity, and 92.0% accuracy. In a previous study by Méndez et al. [51], both time and spectral features from the R-to-R interval (RRi) and QRS area time series were used as inputs to a FFBB ANN aimed at discriminating between apneic and nonapneic segments. A sensitivity of 89%, specificity of 86%, and accuracy of 88% were reached in a minute-by-minute classification, whereas 100% accuracy was achieved when the whole recording is classified as normal or apneic. In a recent study, Nguyen et al. [52] proposed a binary ANN to differentiate apnea from normal sleep based on a hear rate complexity measure by means of the recurrence quantification analysis of HRV recordings. In addition, a SVM classifier and an ensemble combining the decisions from both binary classifiers by means of a confidence score (the weighted sum of the output scores of all binary classifiers) were also assessed. The ensemble reached the highest performance (86.37% sensitivity, 83.47% specificity, 85.26% accuracy), whereas single ANN (85.57% sensitivity, 79.09% specificity, 83.23% accuracy) and the SVM (93.72% sensitivity, 65.88% specificity, 84.14% accuracy) classifiers reached slightly lower accuracy but with an unbalanced sensitivity-specificity pair.

ANNs have been also involved in the detection and characterization of snoring and its reliability in SAHS diagnosis. In the study by Fiz et al. [53], a total of 22 features from time and frequency domains (number of snore episodes, average intensity, and power spectral density parameters) were used as inputs to a MLP ANN. A sensitivity of 87% and a specificity of 71%

were achieved using a SAHS cutoff of 5 e/h, whereas 80% sensitivity and 90% specificity were reached for a cutoff of 15 e/h. In a recent study, Nguyen and Won [54] proposed a novel correlational filter ANN (f-MLP) to distinguish normal breathing patterns from snoring patterns during sleep. This ANN implements a correlational filter operation in the frequency domain in a first hidden layer aimed at improving the discriminant power of the spectral content of input patterns, followed by a second feedforward hidden layer. In this study, the authors reported that the f-MLP classifier reached an average accuracy of 96%, outperforming the conventional MLP approach (82% average accuracy).

EEG signals from nocturnal PSG and ANNs have been also used to detect SAHS. Tagluk et al. [55] estimated the quadratic phase coupling of EEG (C3-A2) using bispectral analysis and trained a MLP ANN to detect patients with SAHS. An overall diagnostic accuracy of 96.15% was reached. In the study by Liu et al. [31], both the EEG energy in the theta band and the pupil size were used as inputs to an ANN aimed at discriminating between SAHS patients and healthy subjects. The authors reported 91% overall accuracy in the classification of both groups. Similarly, in the study by Lin et al. [28], the EEG (C3-O1) signal power in the common frequency bands delta, theta, alpha, and beta were estimated by means of the discrete wavelet transform (DWT) and subsequently used to train a FFBB ANN in order to identify SAHS episodes. A sensitivity of 69.64% and a specificity of 44.44% were obtained. The EEG signal has been also used to classify apnea events into obstructive or central. Akşahin et al. computed the synchronization (coherence and mutual information) between EEG channels (C4-A1 and C3-A2) and fed three different ANN-based binary classifiers: conventional FFBB, RBF, and distributed time-delay (DTD) ANNs [56]. The conventional FFBB ANN reached the highest performance in terms of the mean relative absolute error (MRAE = 0.145).

Features from both thoracic and abdominal effort signals have been also used to classify sleep apneas into obstructive, central, and mixed by means of ANNs. In the study by Fontela-Romero et al. [57], the wavelet coefficients from the DWT of the thoracic effort signal feed a Bayesian feedforward ANN, which achieved a mean accuracy of 83.78 ± 1.90%. Similarly, Tagluk et al. [58] analyzed the abdominal respiration signal by means of the wavelet transform and fed a FFBB ANN aimed at classifying apneic events into obstructive, central, and mixed. The proposed methodology achieved an overall accuracy of 78.5% (obstructive: 73.42%; central: 94.23%; mixed: 66.16%). In a recent study by Guijarro-Berdiñas et al. [59], the thoracic effort signal was used to reach the same goal. The DWT was applied to analyze the frequency content of the signal. The wavelet coefficients compose the input patterns of an ensemble of ANNs, which achieved an overall accuracy of 90.27 ± 0.79% (obstructive: 94.62%; central: 95.47%; mixed: 90.45%).

In the study by Weinreich et al. [60], the spectral entropy was used to analyze the frequency content of airflow recordings and feed an ANN trained to discern among SAHS, Cheyne-Stokes respiration, and normal breathing. An overall accuracy of 91.5% was reached in the classification of airflow patterns into obstructive apneas, periodic respiration, and normal breathing during non-REM sleep. Similarly, Várady et al. [61] trained a feedforward ANN to detect apneic events using respiratory signals. Data from both airflow and respiratory induc-

tance plethysmography were used as inputs to the ANN. Up to 93% of input respiratory patterns were correctly classified into normal, apnea, or hypopnea, although no validation was performed.

ANNs have been also used to combine features from different biomedical recordings. In the study by Belal et al. [62], the correlation coefficients between the heart rate (HR), respiratory rate (RR), and SpO_2 signals were computed to detect apnea events in preterm infants in real time. Principal component analysis (PCA) was applied to the correlation coefficients and the components accounting for the 70% of the total variance of the input data fed the MLP ANN, yielding 81.85% sensitivity, 75.83% specificity, and 76.78% accuracy.

It is noteworthy that most studies in the context of SAHS use ANNs for classification purposes, whereas only a few studies apply regression ANNs to estimate the AHI. This is a more challenging task but also a more useful approach, since the AHI is currently a standardized parameter widely used by physicians to assess SAHS severity and to decide whether the CPAP treatment could be effective. In the aforementioned study by El-Solh et al. [19], the authors compared the agreement of two automated regression approaches with the actual AHI from PSG. Multiple linear regression (MLR) and a regression MLP ANN, both trained with anthropometric and clinical variables, were assessed. Significantly higher correlation was reached using the MLP ANN (0.852 vs. 0.509). In the same way, Marcos et al. [63] used spectral and nonlinear features from nocturnal SpO_2 recordings to feed a regression MLP ANN. High intraclass correlation coefficient was reported (ICC = 0.91), which outperformed the conventional MLR approach (ICC = 0.80). Similarly, in a recent study by Gutiérrez-Tobal et al. [25], regression MLP and RBF ANNs were trained to estimate the AHI from PSG using statistical, spectral, and nonlinear features derived from the airflow signal (thermistor). The estimated AHI from the MLP network reached the highest agreement with the PSG-derived AHI (ICC = 0.849 ± 0.002), improving both the RBF and the conventional MLR models.

A snore-based approach has been proposed by de Silva et al. [64] in order to estimate the actual AHI from PSG. Features from the automated analysis of snoring recordings (pitch, first formant, and the quantified recurrence probability density entropy) and the neck circumference were used as inputs to a FFBB ANN to predict the AHI. Averaged 91 ± 6% sensitivity and 89 ± 5% specificity were obtained using a cutoff of 15 e/h for positive SAHS, whereas for a cutoff of 30 e/h, 86 ± 9% sensitivity and 88 ± 5% specificity were achieved. In a similar subsequent study, de Silva et al. [65] proposed this methodology to characterize differences in snoring sounds due to gender and assessed its influence on the performance of a snore-based SAHS screening model. Using an output threshold of 15 e/h, the gender-dependent regression ANN resulted in increased sensitivity (up to 7% higher) and specificity (up to 6% higher) values compared with the gender-neutral model. In the study by Emoto et al. [66], a MLP ANN was used to predict the current value of the breathing sound signal using the preceding samples, i.e., the target output is the current sample, whereas the d-dimensional input feature pattern is composed by the preceding d samples of the breathing signal. In this way, the ANN was applied to distinguish snoring events from normal breathing comparing the network output with an optimized threshold. The proposed method reached an average sensitivity and specificity values of 89.2 and 87.4%, respectively.

3.2. Automated analysis of PSG: sleep staging and sleep/wake automated detection

In order to identify and quantify the number of respiratory events per hour of sleep and derive the AHI, several neuromuscular and cardiorespiratory recordings from the overnight PSG have to be analyzed. However, the interpretation of a PSG is a complex and laborious task even for trained personnel. In this regard, ANNs have demonstrated to be reliable as well as accurate tools to analyze both the macrostructure (automated sleep staging) and the microstructure (transient pattern detection) of sleep [67]. In the context of sleep staging, nonlinear dynamic measures from EEG in combination with pattern classification algorithms have demonstrated to reach clinically significant results in sleep disorders diagnosis, treatment monitoring, and drug efficacy assessment [68]. In fact, a number of automated algorithms are currently implemented into commercialized software tools for PSG analysis. Nevertheless, the performance of automated pattern recognition algorithms varies greatly depending on the number of stages involved in the classification task, from 2 (wake vs. sleep) to 5 (wake, REM, N1-N3) states (6 classes if the conventional Rechtschaffen and Kales classification is used). In addition, the accuracy is also influenced by the number and kind of recordings involved in the classification task (EEG, EOG, and/or EMG). **Table 2** summarizes the main characteristics of significant studies focused on applications of ANNs for automated sleep staging, arousal quantification, and drowsiness detection.

In the study by Becq et al. [69], the relative power in the common frequency bands of the EEG (C3-A2), as well as the overall variance of EEG and EMG signals, was used to feed a 6-class MLP ANN. The proposed method reached the same performance as a k-NN classifier, achieving 28 ± 2% error rate. Ventouras et al. [70] trained a MLP ANN to detect sleep spindles using a bandpass filtered EEG channel (Cz) without feature extraction. The classifier achieved 80.2% sensitivity and 95.0% specificity in the whole sleep record after a consensus agreement among independent scorers. In the study by Caffarel et al. [71], an ANN-based commercial software using a single-channel EEG (Cz-A1) was assessed. The overall agreement between automated and manual scoring was relatively low in a 4-class classification task ($kappa$ = 0.305) and slightly better in a 2-class classification task ($kappa$ = 0.449). In a later study by Ebrahimi et al. [72], wavelet decomposition and ANNs were used to perform 4-class sleep staging using the EEG signal. An overall sensitivity of 84.2 ± 3.9%, specificity of 94.4 ± 4.5%, and accuracy of 93.0 ± 4.0% were reported. Wavelet coefficients from the EEG (P3-P4) and a backpropagation ANN were also used in the study carried out by Sinha [73]. The author reported accuracies of 96.84%, 93.68%, and 95.52% in the detection of sleep spindles, REM sleep, and awake state, respectively. More recently, Hsu et al. [32] computed energy-based measures from a single EEG channel (Fpz-Cz) to feed a recurrent neural classifier (RNN), which achieved an overall accuracy of 87.2% in a 5-class classification task.

Adding features from additional biomedical signals as inputs to the ANN does not seem to improve significantly the classification performance. In the study by Shambroom et al. [74], a commercial wireless device for automated sleep staging based on the combined activity of EEG, EOG, and EMG is assessed. The Zeo device implements an ANN that achieved 81.1% agreement for light sleep versus deep sleep classification and 93.6% agreement for sleep versus wake classification when the gold standard is a consensus between two independent

Author (year) [Ref.]	ANN model	Purpose	Target function/ class(es)	Input variables	Performance metrics
Becq et al. (2005) [69]	MLP	6-class classification	Wake/NREM 1-4/ REM	EEG (C3-A2) (overall variance, relative power) EMG (overall variance)	ER: 28 ± 2%
Ventouras et al. (2005) [70]	MLP	Binary classification	Sleep spindle detection	Single channel EEG (Cz)	80.2% Se, 95.0% Sp
Caffarel et al. (2006) [71]	NS	4-class 2-class	Wake/light sleep/ deep sleep/REM Wake vs. sleep	EEG (Cz-A1)	$k = 0.305$ $k = 0.449$
Ebrahimi et al. (2008) [72]	NS	4-class classification	Wake/ NREM1+REM/ NREM2/SWS	Wavelet decomposition of single channel EEG	84.2% Se, 94.4% Sp, 93.0% Acc
Sinha (2008) [73]	FFBB	3-class classification	Sleep spindles (SS)/REM/Awake	Wavelet coefficients from EEG	95.35% Acc (overall) 96.84% Acc (SS) 93.68% Acc (REM) 95.52% Acc (Awake)
Hsu et al. (2013) [32]	RNN FFBB PNN	5-class classification	Wake/NREM1/ NREM2/SWS/ REM	Energy features from single-channel EEG	87.2% overall Acc 81.1% overall Acc 81.8% overall Acc
Shambroom et al. (2012) [74]	NS	Binary classification	Sleep *vs.* wake Light *vs.* deed sleep	Combined EEG/EOG/ EMG activity by a single lead (wireless Zeo)	93.6% Acc 81.1% Acc
Griessenberger et al. (2013) [75]	NS	Classification (4-class)	Wake/REM/light sleep/deep sleep	Combined EEG/ EOG/EMG activity by a single lead (wireless Zeo)	72.6% overall Acc
Tagluk et al. (2010) [76]	FFBB	Classification (5-class)	NREM 1 to 4/ REM	Filtered EOG and EMG	74.7% overall Acc 72.6% Acc (NREM1) 73.3% Acc (NREM2) 78.0% Acc (NREM3) 72.3% Acc (NREM4) 77.3% Acc (REM)
Chapotot and Becq (2009) [77]	Ensemble MLP	Classification (6-class)	Wake/N1 to N3/ REM/Movement	Statistical, spectral, nonlinear features from EEG and EMG	36 ± 15% error rate 0.48 ± 0.18 k 34% Acc (Wake) 43% Acc (N1) 51% Acc (N2) 82% Acc (N3) 82% Acc (REM) 13% Acc (Mov.)
Charbonnier et al. (2011) [78]	Ensemble MLP	Classification (5-class)	Wake/NREM1/ NREM2/SWS/ REM	Time and spectral (Fourier analysis) features from EEG, EMG, and EOG	85.5% overall Acc 78.1% Acc (Wake) 64.8% Acc (NREM1) 86.9% Acc (NREM2) 94.8% Acc (SWS) 79.3% Acc (REM)

Author (year) [Ref.]	ANN model	Purpose	Target function/ class(es)	Input variables	Performance metrics
Álvarez-Estevez and Moret-Bonillo (2009) [79]	FLD QD SVM FFBB	Classification (binary)	Arousal detection	Energy in common bands of EEG (Fourier analysis)	0.196 ± 0.015 ER 0.195 ± 0.015 ER 0.140 ± 0.012 ER 0.092 ± 0.010 ER
Patel et al. (2011) [85]	FFBB	Classification (binary)	Alert *vs.* fatigue	Spectral power (Fourier analysis) of HRV	90% Acc
Lin et al. (2006) [86]	FNN	Regression	Driver's drowsiness level estimation	Spectral power of EEG and ICA	Pearson correlation: 0.913 ± 0.027
Kurt et al. (2009) [87]	MLP	Classification (3-class)	Awake/drowsy/ sleep	Wavelet decomposition of EEG, EOG and chin-EMG	97–98% overall Acc
Garcés et al. (2014) [88]	FFBB	Classification (binary)	Alert *vs.* drowsiness	Time, spectral, and wavelet decomposition of single-lead EEG	87.4% Se, 83.6% Sp

Notes: ER: error rate; Se: sensitivity; Sp: specificity; *k*: *kappa* coefficient of classification ability; Acc: accuracy; NS: not specified; MLP: multilayer perceptron; SWS: slow wave sleep; FFBB: feed-forward back-propagation; SS: sleep spindles; RNN: recurrent ANN; PNN: probabilistic neural network; FLD: Fisher's linear discriminant; QD: quadratic discriminant; SVM: support vector machine; FNN: fuzzy ANN; ICA: independent component analysis.

Table 2. Performance and the most relevant characteristics of the studies using ANNs in the context of sleep staging, arousal detection, and drowsiness monitoring.

expert scorers. In a subsequent similar study [75], the same wireless system achieved an over-all agreement of 72.6% in a 4-class approach (Wake, REM, light, and deep sleep). Tagluk et al. [76] used bandpass filtered EOG and EMG recordings as inputs to a feedforward ANN in a 5-class classification task, achieving an overall accuracy of 74.7 ± 1.63%. Similarly, using statistical, spectral, and nonlinear features from EEG and EMG signals and an ensemble clas-sifier based on multiple MLP ANNs, 64% overall performance was achieved for wakefulness, movement, and intermediate sleep detection, while 82% accuracy was reached for deep and paradoxical sleep detection [77]. In the study carried out by Charbonnier et al. [78], 85.5% overall accuracy was reached using EEG-, EOG-, and EMG-derived features as inputs to an ensemble of 4 MLP ANN for 5-class automated sleep staging.

In the study by Álvarez-Estevez and Moret-Bonillo [79], two EEG channels (C2-A2 and C4-A1) and the submental EMG channel were analyzed to automatically detect arousals in the con-text of SAHS classification. For these signals, the energy in the conventional frequency bands was computed by means of the Fourier transform and four automated expert systems were trained: Fisher's linear and quadratic discriminants, a SVM, and a feedforward ANN. The ANN reached the highest performance, achieving 92% accuracy and 0.0921 ± 0.0098 error rate.

Besides ANNs, it is noteworthy that several competing algorithms have been applied for automated sleep staging, such as Gaussian mixture models (88.4% overall accuracy, 6-class,

EEG-based) [80], discrete hidden Markov models (85.29% overall accuracy, 5-class, EEG/EOG/EMG-based) [81], linear (73.7% overall accuracy, 4-class, HRV-based), and quadratic (63.7% overall accuracy, 4-class, HRV-based) discriminant analysis (81% accuracy, 5-class, EEG/EOG/EMG/ECG-based) [82, 83], SVM (89.39% accuracy 5-class, single EEG) [84], DTs (72.6% accuracy, 5-class, EEG/EOG single lead). In the same way as ANNs, these approaches are characterized by variable performance.

3.2.1. Driver's drowsiness detection

A relevant application of ANNs in the context of SAHS is the detection of drivers' fatigue and/or drowsiness, which is an important issue for patients suffering from SBD. In this regard, different physiological signals have been used to monitor alertness, such as spectral analysis of HRV (90% accuracy) [85] and EEG (0.913 ± 0.027 correlation between actual and estimated alertness levels) [86], wavelet coefficients of EEG combined with features from EOG and EMG (97–98% 3-class overall accuracy) [87], and time, spectral, and wavelet features from single-lead EEG [88]. Neuromuscular (EEG, EOG, EMG) and cardiac (ECG) signals have been analyzed predominantly in order to detect drowsiness, though additional physiological recordings (oximetry, skin conductance), physical measures (eye movement/blinks, face and mouth images), and driver's performance measures (steering wheel movements) have been also proposed as inputs to different pattern recognition methods, specially Bayesian networks, SVMs, and ensembles of linear classifiers [89–91]. The main limitation of these automated algorithms is that a great amount of data is needed to perform an accurate training of the pattern recognition method. Nonetheless, alertness monitoring systems are already incorporated in many high-end vehicles.

3.3. Neural networks and continuous positive airway pressure

The incorporation of automated decision support systems in the common clinical practice of SAHS diagnosis is still very limited. Conversely, the implementation of artificial intelligence-based expert systems in treatment devices for sleep-related breathing disorders therapy increased significantly during the last decade. In this regard, the exponential technological development of continuous positive airway pressure (CPAP) devices relies on the automated analysis of breathing patterns by means of expert systems, most of them based on ANNs. Currently, CPAP is the primary preferred treatment of mild, moderate, and severe SAHS and thus it is considered the standard of care. During CPAP treatment, a continuous pressure of air is delivered to the patient's upper airway to keep patency [92]. Though nonintrusive, simple, and effective, the device delivers an unnecessary constant high pressure during the whole night whatever the actual patient's needs, which decreases comfort and in turn treatment compliance. This is the main limitation of CPAP and thus the most relevant improvements during the last years focused on the modulation of the pressure delivered by the device in order to fit patient's needs. In this regard, the major companies operating in the SAHS therapy market incorporated to their devices automated algorithms to monitor and modulate the breathing gas pressure. Nevertheless, most manufactures provide no technical data about the design and implementation of their automated signal processing algorithms and thus they are blackboxes hard to interpret and assess.

As aforementioned, determining the optimal therapeutic pressure has been a major goal of research regarding CPAP treatment. Different respiratory-related signals have been assessed for automated regulation of the pressure. Airflow, SpO_2 from oximetry, HRV, pharyngeal wall vibration, and snoring sounds have been involved in automated algorithms aimed at detecting airflow limitation and respiratory events. Among them, the analysis of the airflow profile is the most widely used method [93, 94]. In this regard, several algorithms have been patented during the last years, which reflect the increasing interest of leading companies in this field. In the patent by Norman et al. [93], a pretrained ANN fed with shape-based features from the airflow signal is used to detect the presence of airflow limitation in each individual patient's breath. Eklund et al. [95] granted a patent for automatically adjusting the flow pressure when respiratory events are detected. To achieve this goal, an ANN is fed with respiration-related variables. In a recent granted patent, Waxman et al. [96] proposed a Large Memory Storage and Retrieval (LAMSTAR) neural network to process patient's physiological data in order to predict breathing events and control the airway pressure level supplied to the user. This algorithm reached high prediction ability within the 30 s preceding the respiratory event [96]. Similarly, in the patent by Hedner et al. [97], the authors describe a pattern recognition system based on a plurality of ANNs aimed at controlling the therapy breathing support in order to increase its effectiveness. Leading companies, such as Philips Respironics, ResMed, or Fisher & Paykel, incorporated these algorithms into their CPAP devices. Nevertheless, additional research is still needed to further assess whether these technological advances can effectively improve CPAP adherence.

Automatic detection of wake and sleep states is a novel approach for enhancing patient's comfort [98, 99]. In the study carried out by Ayappa et al. [98], the authors proposed an ANN to detect irregular respiration characteristics of sleep/wake transitions. In this study, the CPAP flow signal is parameterized by means of breath timing and amplitude measures, which subsequently feed the ANN in order to detect irregular breathing. This algorithm is used in the commercial system SensAwake™ (Fisher & Paykel, Auckland, NZ) in order to automatically decrease the therapeutic pressure when the patient is awake [99]. This ANN has demonstrated to be effective for sleep onset and awakening detection, though there is still little if any evidence supporting its actual long-term influence on patient's comfort and CPAP compliance.

In order to obtain the optimal CPAP pressure level for a patient, an individual titration procedure is needed. This technique is aimed at estimating the continuous pressure that normalizes the patient's sleep and breathing during in-lab PSG, which contributes to increase the large waiting lists. Therefore, alternative methods are demanded. In this regard, El-Solh et al. [100] designed and trained a GRNN aimed at estimating the most effective continuous pressure using demographic and anthropometric variables (those from the Hoffstein formula, i.e., age, gender, BMI, neck circumference, and AHI). The authors reported high agreement between the optimal pressure determined by standard titration during overnight PSG and the pressure predicted by the ANN. In a later randomized study, El-Solh et al. [101] reported that this ANN can be effectively used to guide CPAP titration. The authors showed that automated titration procedures using this methodology reached the optimal CPAP pressure at a shorter time interval compared to conventional PSG-based titration, as well as lower titration failure.

4. Conclusion

Researchers carried out an exhaustive study during the last decades focused on the design of automated expert systems derived from artificial intelligence able to help physicians in their daily practice. Accordingly, several computer-aided decision support systems have been proposed to overcome limitations of the standard diagnostic methodology for SAHS. Among all the automated prediction methods, ANNs are probably the most widely used pattern recognition algorithm in the context of SAHS management. Their flexibility to model complex nonlinear problems and their higher generalization ability allow ANNs to reach higher performance rates both in classification and regression problems. In this regard, several applications of ANNs have been developed, such as classification of patients suspected of suffering from SAHS, AHI estimation, detection and quantification of respiratory events, apneic events classification, automated sleep staging and arousal detection, alertness monitoring systems, and airflow pressure optimization in PAP-based devices. On the other hand, the most common limitation of ANNs relates to the interpretation of the results in terms of the significance of the variables involved in the model. In this way, ANNs are most times viewed as blackboxes that are not able to generate understandable rules, which is the main weakness of neural-based classifiers. Conversely, both decision trees and probabilistic networks also reach high performance by providing interpretable rules and relationships between input variables.

Regarding input features, ANNs are able to deal with high-dimensional spaces composed of several features. This is especially useful when working with a lot of data sources providing information about the problem under study, such as symptoms reported by the patient, physical examination, sleep questionnaires, or PSG, among others. However, it is important to highlight that, sometimes, researchers try to compose a wide initial feature set in order to gather as much information as possible, including features from signal processing algorithms regardless of their relevance or clinical meaning. In this way, feature selection strategies are very useful to distinguish the more significant ones. In addition, dimensionality reduction algorithms allow ANNs to deal with the curse of dimensionality problem and to control over-fitting. Nevertheless, just a few studies apply feature selection techniques before the classification stage.

ANNs have yielded reliable and accurate applications in the context of SAHS detection. Nevertheless, it is noteworthy that, in the last years, there is a trend to use different pattern recognition algorithms, particularly SVMs and ensemble classifiers. SVMs have emerged as powerful tools able to achieve significantly high performance both in classification and regression problems. They are kernel-based maximum margin classifiers, i.e., the decision boundary is determined by a subset of the training data samples in a transformed space in which the margin (the distance between the boundary and the closest samples) is maximized. In this way, the optimization problem is relatively straightforward [18]. Several recent studies have demonstrated the usefulness of SVMs in the framework of SAHS management [102–105]. Moreover, in the present research, some studies were reviewed reporting that SVM-based

classifiers reached higher accuracy than ANNs [23, 45, 52]. Unlike ANNs, SVMs are capable to minimize both structural and empirical risk, leading to higher generalization ability even when working with limited training datasets [103]. On the other hand, they are also characterized as blackboxes and usually higher computational time is needed to optimize the classifier [27]. Unfortunately, there are few studies assessing the performance of different classification approaches in the same conditions (population under study and equal optimization of input parameters), leading to biased results and poor generalization. Open access databases, such as the *Physionet* or the *Sleep Heart Health Study* (SHHS), provide a common benchmark to properly assess the performance of different methodologies using the same data. Nevertheless, these databases are limited and most studies are carried out using no publicly available datasets, which restricts comparisons.

In addition, it is also noteworthy that ensemble classifiers, from the simplest majority vote to the more complex bagging, boosting, and stacking algorithms, have been recently introduced in the context of SAHS in order to improve classification performance [106, 107]. It is obvious that misclassified samples are not always the same when using different classification algorithms. Accordingly, improved performance may be reached when working with several classifiers at the same time. In this way, ensemble algorithms take advantage of the information provided by all the classifiers involved in the classification or regression task. The studies by Guijarro-Berdiñas et al. [59] and Nguyen et al. [52] demonstrated the reliability and efficacy of ANN-based ensembles. Nevertheless, further research is still need in order to exploit the full potential of this approach in the context of SAHS diagnosis.

Acknowledgements

Author details

Daniel Álvarez[1, 2], Ana Cerezo-Hernández[1], Graciela López-Muñiz[1], Tania Álvaro-De Castro[1], Tomás Ruiz-Albi[1], Roberto Hornero[2] and Félix del Campo[1, 2]*

*Address all correspondence to: fsas@telefonica.net

1 Sleep-Ventilation Unit, Pneumology Service, Río Hortega University Hospital, Valladolid, Spain

2 Biomedical Engineering Group, University of Valladolid, Valladolid, Spain

References

[1] Lisboa PJ, Taktak AFG. The use of artificial neural network in decision supports in cancer. A systematic review. Neural Networks. 2006;19:408–415. DOI: 10.1016/j.neunet.2005.10.007

[2] Adams ST, Leveson SH. Clinical prediction rules. BMJ. 2012;**344**:d8312–d8312. DOI: 10.1136/bmj.d8312

[3] Marcos JV, Hornero R, Álvarez D, Del Campo F, Aboy M. Automated detection of obstructive sleep apnoea syndrome from oxygen saturation recordings using linear discriminant analysis. Medical & Biological Engineering & Computing. 2010;**48**:895–902. DOI: 10.1007/s11517-010-0646-6

[4] Álvarez D, Hornero R, Marcos JV, Del Campo F. Feature selection from nocturnal oximetry using genetic algorithms to assist in obstructive sleep apnoea diagnosis. Medical Engineering & Physics. 2012;**34**:1049–1057. DOI: 10.1016/j.medengphy.2011.11.009

[5] Pandey B, Mishra RB. Knowledge and intelligent computing system in medicine. Computers in Biology and Medicine. 2009;**39**:215-30. DOI: 10.1016/j.compbiomed.2008.12.008

[6] Raudys S. Statistical and neural classifiers. An integrated approach to design. London: Springer; 2001. ISBN: 1-85233-297-2

[7] McCulloch WS, Pitts W. A logical calculus of the ideas immanent in nervous activity. Bulletin of Mathematical Biophysics. 1943;**5**:115. DOI:10.1007/BF02478259

[8] Flemons WW, Littner MR, Rowlet JA, Gay P, Anderson WM, Hudgel DW, McEvoy RD, Loube DI. Home diagnosis of sleep apnea: A systematic review of the literature. Chest. 2003;**124**:1543–1579. DOI: 10.1378/chest.124.4.1543

[9] Qaseem A, Dallas P, Owens DK, Starkey M, Holty J-EC, Shekelle P. Diagnosis of obstructive sleep apnea in adults: A clinical practice guideline from the American College of Physicians. Annals of Internal Medicine. 2014;**161**:210–220. DOI: 10.7326/M12-3187

[10] Álvarez-Estevez D, Moret-Bonillo V. Computer-assisted diagnosis of the sleep apnea-hypopnea syndrome: a review. Sleep Disorders. 2015;**2015**:1–33. DOI: 10.1155/2015/237878

[11] Kocak O, Bayrak T, Erdamar A, Ozparlak L, Telatar Z, Erogul O. Automated detection and classification of sleep apnea types using electrocardiogram (ECG) and electroencephalogram (EEG) features. Advances in Electrocardiograms – Clinical Applications, PhD. Richard Millis (Ed.). Croatia: InTech; 2012. ISBN: 978-953-307-902-8.

[12] Álvarez D, Gutiérrez-Tobal GC, Del Campo F, Hornero R. Positive airway pressure and electrical stimulation methods for obstructive sleep apnea treatment: a patent review (2005–2014). Expert Opinion on Therapeutic Patents. 2015;**25**:971–989. DOI: 10.1517/13543776.2015.1054094

[13] Haykin S. Neural networks: A comprehensive foundation. New Jersey: Prentice Hall Inc; 1999.

[14] Rojas R. Neural networks. A systematic introduction. Berlin Heidelberg: Springer Verlag; 1996. DOI: 10.1007/978-3-642-61068-4

[15] Zhang GP. Neural networks for classification: a survey. IEEE Transactions on Systems, Man and Cibernetics, Part C. 2000;**30;**451–462. DOI: 10.1109/5326.897072

[16] Hush DR, Horne BG. Progress in supervised neural networks. IEEE Signal Processing Magazine. 1993;**10:**8–39. DOI: 10.1109/79.180705

[17] Bishop CM. Neural networks for pattern recognition. New York: Oxford University Press; 1995. ISBN: 0-19-853864

[18] Bishop CM. Pattern recognition and machine learning. New York: Springer; 2006. ISBN: 0-387-31073-8

[19] El-Solh AA, Mador MJ, Ten-Brock E, Shucard DW, Abul-Khoudoud M, Grant BJ. Validity of neural network in sleep apnea. Sleep. 1999;**22**:105–111.

[20] Marcos JV, Hornero R, Álvarez D, Del Campo F, Zamarrón C, López M. Utility of mul-tilayer perceptron neural network classifiers in the diagnosis of the obstructive sleep apnoea syndrome from nocturnal oximetry. Computer Methods and Programs in Biomedicine. 2008;**92**:79–89. DOI: 10.1016/j.cmpb.2008.05.006

[21] Karamanli H, Yalcinoz T, Yalcinoz MA, Yalcinoz T. A prediction model based on artificial neural networks for the diagnosis of obstructive sleep apnea. Sleep Breath. 2016;**20**:509–514. DOI 10.1007/s11325-015-1218-7

[22] Marcos JV, Hornero R, Álvarez D, Nabney IT, Del Campo F, Zamarrón C. The classifi-cation of oximetry signals using Bayesian neural networks to assist in the detection of obstructive sleep apnoea syndrome. Physiological Measurement. 2010;**31**:375–394. DOI: doi:10.1088/0967-3334/31/3/007

[23] Khandoker AH, Karmakar CHK, Palaniswami M. Automated recognition of patients with obstructive sleep apnoea using wavelet-based features of electrocardiogram recordings. Computers in Biology and Medicine. 2009;**39**:88–96. DOI: 10.1016/j.compbiomed.2008.11.003

[24] Marcos JV, Hornero R, Álvarez D, Del Campo F, López M, Zamarrón C. Radial basis function classifiers to help in the diagnosis of the obstructive sleep apnoea syndrome from nocturnal oximetry. Medical & Biological Engineering & Computing. 2008;**46**:323–332. DOI: 10.1007/s11517-007-0280-0

[25] Gutiérrez-Tobal GC, Álvarez D, Marcos JV, Del Campo F, Hornero R. Pattern recogni-tion in airflow recordings to assist in the sleep apnoea–hypopnoea syndrome diagno-sis. Medical & Biological Engineering & Computing. 2013;**51**:1367–1380. DOI 10.1007/s11517-013-1109-7

[26] Huang S-H, Teng N-C, Wang K-J, Chen K-H, Lee H-C, Wang P-C. Use of oximetry as a screening tool for obstructive sleep apnea: a case study in Taiwan. Journal of Medical Systems. 2015;**39**:29. DOI 10.1007/s10916-015-0195-5

[27] Wang K-J, Chen K-H, Huang S-H, Teng N-C. A prognosis tool based on fuzzy anthropometric and questionnaire data for obstructive sleep apnea. Journal of Medical Systems. 2016;**40**:110. DOI: 10.1007/s10916-016-0464-y

[28] Lin R, Lee R-G, Tseng C-L, Zhou H-K, Chao C-F, Jiang J-A. A new approach for identifying sleep apnea syndrome using wavelet transform and neural networks. Biomedical Engineering: Applications, Basis and Communications. 2006;**18**:138–143. DOI: 10.4015/S1016237206000233

[29] Polat K, Yosunkaya S, Güneş S. Comparison of different classifier algorithms on the automated detection of obstructive sleep apnea syndrome. Journal of Medical Systems. 2008;**32**:243–250. DOI 10.1007/s10916-008-9129-9

[30] Ghandeharioun H, Rezaeitalab F, Lotfi R. Accurate estimation of obstructive sleep apnea severity using non-polysomnographic features for home-based screening. Iranian Journal of Public Health. 2015;**44:**1433–1435.

[31] Liu D, Pang Z, Lloyd SR. A neural network method for detection of obstructive sleep apnea and narcolepsy based on pupil size and EEG. IEEE Transactions on Neural Networks. 2008;**19**:308–318. DOI: 10.1109/TNN.2007.908634

[32] Hsu Y-L, Yang Y-T, Wang J-S, Hsu C-Y. Automatic sleep stage recurrent neural classifier using energy features of EEG signals. Neurocomputing. 2013;**104:**105–114. DOI: 10.1016/j.neucom.2012.11.003

[33] Hornik K. Approximation capabilities of multilayer feedforward networks. Neural Networks. 1991;**4:**251–257. DOI: 10.1016/0893-6080(91)90009-T

[34] Basheer IA, Hajmeer M. Artificial neural networks: fundamentals, computing, design, and application. Journal of Microbiological Methods. 2000;**43**:3–31. DOI: 10.1016/S0167-7012(00)00201-3

[35] Sethi IK, Jain AK. Artificial neural networks and statistical pattern recognition. Machine intelligence and pattern recognition series, VOL. II. Amsterdam: Elsevier; 1991. ISBN: 0-444-88740-7

[36] Jain AK, Duin RPW, Mao J. Statistical pattern recognition: a review. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2000;**22;**4–37. DOI: 10.1109/34.824819

[37] Foster KR, Koprowski R, Skufca JD. Machine learning, medical diagnosis, and biomedical engineering research – commentary. BioMedical Engineering OnLine. 2014;**13**:94. DOI: 10.1186/1475-925X-13-94

[38] Dayhoff JE, Deleo JM. Artificial neural networks: opening the black box. Cancer. 2001;**91**:1615–1635.

[39] Graupe D. Principles of artificial neural networks. 3rd edition. Singapore: World Scientific Publishing Co.; 2007. ISBN: 978-981-4522-73-1

[40] Baxt WG. Application of artificial neural networks to clinical medicine. Lancet. 1995;**346**:1135–1138.

[41] Siristatidis CS, Chrelias C, Pouliakis A, Katsimanis E, Kassanos D. Artificial neural networks in gynaecological diseases: current and potential future applications. Medical Science Monitor. 2010;**16**:RA231-6.

[42] Amato F, López A, Peña-Méndez EM, Vaňhara P, Hampl A, Havel J. Artificial neural networks in medical diagnosis. Journal of Applied Biomedicine. 2013;**11**:47–58. DOI:10.2478/v10136-012-0031-x

[43] Kirby SD, Danter W, George CFP, Francovic T, Ruby RRF, Ferguson KA. Neural network prediction of obstructive sleep apnea from clinical criteria. Chest. 1999;**116**:409–415. DOI: 10.1378/chest.116.2.409

[44] De Chazal P, Heneghan C, McNicholas WT. Multimodal detection of sleep apnoea using electrocardiogram and oximetry signals. Philosophical Transactions of the Royal Society A. 2009;**367**:369–389. DOI: 10.1098/rsta.2008.0156

[45] Su CT, Chen KH, Chen LF, Wang PC, Hsiao YH. Prediagnosis of obstructive sleep apnea via multiclass MTS. Computational and Mathematical Methods in Medicine. 2012;**2012**:212498. DOI: 10.1155/2012/212498

[46] Almazaydeh L, Faezipour M, Elleithy K. A, editors. Neural network system for detection of obstructive sleep apnea through SpO_2 signal features. International Journal of Advanced Computer Science and Applications. 2012;**3**:7–11. ISBN: 2156-5570

[47] Sánchez-Morillo D, Gross N. Probabilistic neural network approach for the detection of SAHS from overnight pulse oximetry. Medical & Biological Engineering & Computing. 2013;**51**:305. DOI 10.1007/s11517-012-0995-4

[48] Khandoker AH, Gubbi J, Palaniswami M. Recognizing central and obstructive sleep apnea events from normal breathing events in ECG recordings. Computers in Cardiology. 2008;**35**:681–684. DOI: 10.1109/CIC.2008.4749133

[49] Acharya UR, Chua ECP, Faust O, Lim TC, Lim LFB. Automated detection of sleep apnea from electrocardiogram signals using nonlinear parameters. Physiological Measurement. 2011;**32:**287–303. DOI: 10.1088/0967-3334/32/3/002

[50] Lweesy K, Fraiwan L, Khasawnch N, Dickhuash. New automated detection method of OSA based on artificial neural networks using P wave shape and time changes. Journal of Medical Systems. 2011;**35**:723–734. DOI: DOI 10.1007/s10916-009-9409-z

[51] Méndez MO, Bianchi AM, Matteucci M, Cerutti S, Penzel T. Sleep apnea screening by autoregressive models from a single ECG lead. IEEE Transactions on Biomedical Engineering. 2009;**56**:2838–2850. DOI: 10.1109/TBME.2009.2029563

[52] Nguyen HD, Wilkins BA, Cheng Q, Benjamin BA. An online sleep apnea detection method based on recurrence quantification analysis. IEEE Journal of Biomedical and Health Informatics. 2014;**18**:1285–1293] DOI: 10.1109/JBHI.2013.2292928

[53] Fiz JA, Jané R, Solá-Soler J, Abad J, García M, Morera J. Continuous analysis and monitoring of snores and their relationship to the apnea-hypopnea index. Laryngoscope. 2010;**120**:854–862. DOI: 10.1002/lary.20815

[54] Nguyen TL, Won Y. Sleep snoring detection using multi-layer neural networks. Bio-Medical Materials and Engineering. 2015;**26:**S1749–S1755. DOI: 10.3233/BME-151475

[55] Tagluk ME, Segzin N. A new approach for estimation of obstructive sleep apnea syndrome. Expert Systems with Applications. 2011;**38**:5346–5351. DOI: 10.1016/j.eswa.2010.10.022

[56] Akşahin M, Aydın S, Fırat H, Eroğul O. Artificial apnea classification with quantitative sleep EEG synchronization. Journal of Medical Systems 2012;**36**(1):139–144. DOI: 10.1007/s10916-010-9453-8

[57] Fontenla-Romero O, Guijarro-Berdiñas B, Alonso-Betanzos A, Moret-Bonillo V. A new method for sleep apnea classification using wavelets and feedforward neural networks. Artificial Intelligence in Medicine. 2005;**34**:65–76. DOI: 10.1016/j.artmed.2004.07.014

[58] Tagluk ME, Akin M, Sezgin N. Classıfıcation of sleep apnea by using wavelet transform and artificial neural networks. Expert Systems with Applications. 2010;**37**:1600–1607. DOI:

[59] Guijarro-Berdiñas B, Hernández-Pereira E, Peteiro-Barral D. A mixture of experts for classifying sleep apneas. Expert Systems with Applications. 2012;**39:**7084–7092. DOI: 10.1016/j.eswa.2012.01.037

[60] Weinreich G, Armitstead J, Teschler H. Pattern recognition of obstructive sleep apnoea and Cheyne–Stokes respiration. Physiological Measurement. 2008;**29**:869. DOI: 10.1088/0967-3334/29/8/002

[61] Várady P, Micsik T, Benedek S, Benyó Z. A novel method for the detection of apnea and hypopnea events in respiration signals. IEEE Transactions on Biomedical Engineering. 2002;**49**:936–942. DOI: 10.1109/TBME.2002.802009

[62] Belal SY, Emmerson AJ, Beatty PC. Automatic detection of apnoea of prematurity. Physiological Measurement. 2011;**32**(5):523–542. DOI: 10.1088/0967-3334/32/5/003

[63] Marcos JV, Hornero R, Álvarez D, Aboy M, Del Campo F. Automated prediction of the apnea-hypopnea index from nocturnal oximetry recordings. IEEE Transactions on Biomedical Engineering. 2012;**59**:141–149. DOI: 10.1109/TBME.2011.2167971

[64] de Silva S, Abeyratne UR, Hukins C. A method to screen obstructive sleep apnea using multi-variable non-intrusive measurements. Physiological Measurement. 2011;**32**(4):445–465. DOI: 10.1088/0967-3334/32/4/006

[65] de Silva S, Abeyratne UR, Hukins C. Impact of gender on snore-based obstructive sleep apnea screening. Physiological Measurement. 2012;**33**:587. DOI: 10.1088/0967-3334/33/4/587

[66] Emoto T, Abeyratne UR, Chen Y, Kawata I, Akutagawa M, Kinouchi Y. Artificial neural networks for breathing and snoring episode detection in sleep sounds. Physiological Measurement. 2012;**33**(10):1675–1689. DOI: 10.1088/0967-3334/33/10/1675

[67] Motamedi-Fakr S, Moshrefi-Torbati M, Hill M, Hill CM, White PR. Signal processing techniques applied to human sleep EEG signals—A review. Biomedical Signal Processing and Control. 2014;**10**:21–33. DOI: 10.1016/j.bspc.2013.12.003

[68] Acharya UR, Bhat S, Faust O, Adeli H, Chua EC-P, Lim WJE, Koh JEW. Nonlinear dynamics measures for automated EEG-based sleep stage detection. European Neurology. 2015;**74**:268–287. DOI: 10.1159/000441975

[69] Becq G, Charbonnier S, Chapotot F, Buguet A, Bourdon L, Baconnier P. Comparison between five classifiers for automatic scoring of human sleep recordings. Studies in Computational Intelligence (SCI) Vol. 4. Berlin, Heidelberg: Springer-Verlag; 2005. pp. 113–127. DOI: DOI: 10.1007/11011620_8

[70] Ventouras EM, Monoyiou EA, Ktonas PY, Paparrigopoulos T, Dikeos DG, Uzunoglu NK, Soldatos CR. Sleep spindle detection using artificial neural networks trained with filtered time-domain EEG: A feasibility study. Computer Methods and Programs in Biomedicine. 2005;**78**:191–207. DOI: 10.1016/j.cmpb.2005.02.006

[71] Caffarel J, Gibson GJ, Harrison JP, Griffiths CJ, Drinnan MJ. Comparison of manual sleep staging with automated neural network-based analysis in clinical practice. Medical & Biological Engineering & Computing. 2006;**44**:105–110. DOI: 10.1007/s11517-005-0002-4

[72] Ebrahimi F, Mikaeili M, Estrada E, Nazeran H. Automatic sleep stage classification based on EEG signals by using neural networks and wavelet packet coefficients. In: Proceedings of the 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society; 20–25 August 2008; Vacouver. New York: IEEE; 2008. pp. 1151–1154. DOI: 10.1109/IEMBS.2008.4649365

[73] Sinha RK. Artificial neural network and wavelet based automated detection of sleep spindles, REM sleep and wake states. Journal of Medical Systems. 2008;**32**:291–299. DOI: 10.1007/s10916-008-9134-z

[74] Shambroom JR, Fábregas SE, Johnstone J. Validation of an automated wireless system to monitor sleep in healthy adults. Journal of Sleep Research. 2012;**21**:221–230. DOI: 10.1111/j.1365-2869.2011.00944.x

[75] Griessenberger H, Heib DPJ, Kunz AB, Hoedlmoser K, Schabus M. Assessment of a wireless headband for automatic sleep scoring. Sleep Breath. 2013;**17**:747–752. DOI 10.1007/s11325-012-0757-4

[76] Tagluk ME, Sezgin N, Akin M. Estimation of sleep stages by an artificial neural network employing EEG, EMG and EOG. Journal of Medical Systems. 2010;**34**:717–725. DOI 10.1007/s10916-009-9286-5

[77] Chapotot F, Becq G. Automated sleep–wake staging combining robust feature extraction, artificial neural network classification, and flexible decision rules. International Journal of Adaptive Control and Signal Processing. 2009;**1**:1–15. DOI: 10.1002/acs.1147

[78] Charbonnier S, Zoubek L, Lesecq S, Chapotot F. Self-evaluated automatic classifier as a decision-support tool for sleep/wake staging. Computers in Biology and Medicine. 2011;**41**:380–389. DOI:10.1016/j.compbiomed.2011.04.001

[79] Álvarez-Estévez D, Moret-Bonillo V. Model comparison for the detection of EEG arousals in sleep apnea patients. In: Bio-Inspired Systems: Computational and Ambient Intelligence, Volume 5517 of the series Lecture Notes in Computer Science. Berlin, Heidelberg: Springer; 2009. pp. 997–1004. DOI: 10.1007/978-3-642-02478-8_125

[80] Acharya UR, Chua EC-P, Chua KC, Min LC, Tamura T. Analysis and automatic identification of sleep stages using higher order spectra. International Journal of Neural Systems. 2010;**20:**509. DOI: 10.1142/S0129065710002589

[81] Pan S-T, Kuo C-E, Zeng J-H, Liang S-F. A transition-constrained discrete hidden Markov model for automatic sleep staging. BioMedical Engineering OnLine. 2012;**11**:52. DOI: 10.1186/1475-925X-11-52

[82] Ebrahimi F, Setarehdan S-K, Ayala-Moyeda J, Nazeran H. Automatic sleep staging using empirical mode decomposition, discrete wavelet transform, time-domain, and nonlinear dynamics features of heart rate variability signals. Computer Methods and Programs in Biomedicine. 2013;**112**:47–57. DOI: 10.1016/j.cmpb.2013.06.007

[83] Krakovská A, Mezeiová K. Automatic sleep scoring: A search for an optimal combination of measures. Artificial Intelligence in Medicine. 2011;**53:**25–33. DOI: 10.1016/j.artmed.2011.06.004

[84] Koley B, Dey D. An ensemble system for automatic sleep stage classification using single channel EEG signal. Computers in Biology and Medicine. 2012;**42:**1186–1195. DOI: 10.1016/j.compbiomed.2012.09.012

[85] Patel M, Lal SKL, Kavanagh D, Rossiter P. Applying neural network analysis on heart rate variability data to assess driver fatigue. Expert Systems with Applications. 2011;**38:**7235–7242. DOI: 10.1016/j.eswa.2010.12.028

[86] Lin C-T, Ko L-W, Chung I-F, Huang T-Y, Chen Y-C, Jung T-P, Liang S-F. Adaptive EEG-based alertness estimation system by using ICA-based fuzzy neural networks. IEEE Transactions on Circuits and Systems. 2006;**53**:2469–2476. DOI:10.1109/TCSI.2006.884408

[87] Kurt MB, Sezgin N, Akin M, Kirbas G, Bayram M. The ANN-based computing of drowsy level. Expert Systems with Applications. 2009;**36**:2534–2542. DOI: 10.1016/j.eswa.2008.01.085

[88] Garcés A, Orosco L, Laciar E. Automatic detection of drowsiness in EEG records based on multimodal analysis. Medical Engineering & Physics. 2014;**36**:244–249. DOI: 10.1016/j.medengphy.2013.07.011

[89] Dong Y, Hu Z, Uchimura K, Murayama N. Driver inattention monitoring system for intelligent vehicles: A review. IEEE Transactions on Intelligent Transportation Systems. 2011;**12**:596–614. DOI:10.1109/TITS.2010.2092770

[90] Sahayadhas A, Sundaraj K, Murugappan M. Detecting driver drowsiness based on sensors: A review. Sensors. 2012;**12**:16937–16953. DOI:10.3390/s121216937

[91] DababnehRelated L, El-GindyRelated M. Driver vigilance level detection systems: a literature survey. International Journal of Vehicle Performance. 2016;**2**:1–29. DOI: 10.1504/IJVP.2015.074120

[92] Gharibeh T, Mehra R. Obstructive sleep apnea syndrome: natural history, diagnosis and emerging treatment options. Nature and Science of Sleep 2010;**2**:233–55. DOI: 10.2147/NSS.S6844

[93] Norman RG, Rapoport DM, Ayappa I. Detection of flow limitation in obstructive sleep apnea with an artificial neural network. Physiological Measurement. 2007;**28**:1089. DOI: 10.1088/0967-3334/28/9/010

[94] Hertegonne K, Bauters F. The value of auto-adjustable CPAP devices in pressure titration and treatment of patients with obstructive sleep apnea syndrome. Sleep Medicine Reviews. 2010;**14**:115–119. DOI: 10.1016/j.smrv.2009.07.001

[95] Eklund O, Bergfalk H, Hedner JA, Knagenhjelm HP. Auto CPAP. Patent assignee: Breas Medical AB. United States Patent Number: US6889691, 2005.

[96] Waxman J, Graupe D, Carley DW. Detection and prediction of physiological events in people with sleep disordered breathing using a LAMSTAR neural network. Patent assignee: The Board of Trustees of the University of Illinois. United States Patent Number: US8775340, 2014.

[97] Hedner J, Knagenhjelm P, Tiedje M, Grote L. Multilevel ventilator. Patent assignee: Breas Medical AB. European Patent Number: EP1750581, 2014.

[98] Ayappa I, Norman RG, Whiting D, Tsai AHW, Anderson F, Donnely E, Silberstein DJ, Rapoport DM. Irregular respiration as a marker of wakefulness during titration of CPAP. Sleep. 2009;**32**:99–104.

[99] Dungan II GC, Marshall NS, Hoyos CM, Yee BJ, Grunstein RR. A randomized crossover trial of the effect of a novel method of pressure control (SensAwake) in automatic continuous positive airway pressure therapy to treat sleep disordered breathing. Journal of Clinical Sleep Medicine 2011;**7**:261–267. DOI: 10.5664/JCSM.1066

[100] El-Solh AA, Aldik Z, Alnabhan M, Grant B. Predicting effective continuous positive airway pressure in sleep apnea using an artificial neural network. Sleep Medicine. 2007;**8**:471–477.

[101] El-Solh AA, Akinnusi M, Patel A, Bhat A, Tenbrock R. Predicting optimal CPAP by neural network reduces titration failure: a randomized study. Sleep & Breathing. 2009;**13**:325–330. DOI: 10.1007/s11325-009-0247-5

[102] Álvarez D, Hornero R, Marcos JV, Wessel N, Penzel T, Glos M, Del Campo F. Assessment of feature selection and classification approaches to enhance information from overnight oximetry in the context of sleep apnea diagnosis. International Journal of Neural Systems. 2013;**23**:1–18. DOI: 10.1142/S0129065713500202

[103] Al-Angari HM, Sahakian AV. Automated recognition of obstructive sleep apnea syndrome using support vector machine classifier. IEEE Transactions on Information Technology in Biomedicine. 2012;**16**:463–468. DOI: 10.1109/TITB.2012.2185809

[104] Fekr AR, Janidarmian M, Radecka K, Zilic Z. A medical cloud-based platform for respiration rate measurement and hierarchical classification of breath disorders. Sensors. 2014;**14**:11204–11224. DOI: 10.3390/s140611204

[105] Lee BG, Lee BL, Chung WY. Mobile healthcare for automatic driving sleep-onset detection using wavelet-based EEG and respiration signals. Sensors. 2014;**14**:17915–17936. DOI: 10.3390/s141017915

[106] Xie B, Minn H. Real-time sleep apnea detection by classifier combination. IEEE Transactions on Information Technology in Biomedicine. 2012;**16**:469–477. DOI: 10.1109/TITB.2012.2188299

[107] Gutiérrez-Tobal GC, Álvarez D, Del Campo F, Hornero R. Utility of AdaBoost to detect sleep apnea-hypopnea syndrome from single-channel airflow. IEEE Transactions on Biomedical Engineering. 2016;**63**:636–646. DOI: 10.1109/TBME.2015.2467188