Diagnosing Chronic Obstructive Pulmonary Disease with Artificial Neural Networks using Health Expert Guidelines

Maria Angélica de Oliveira Camargo-Brunetto[1] and André R. Gonçalves[2]

[1]Computer Science Department, State University of Londrina, Londrina, Paraná, Brazil

[2]School of Electrical and Computer Engineering, University of Campinas, Campinas, São Paulo, Brazil

Keywords: Decision Support System, Artificial Neural Networks, COPD, Multilayer Perceptron, Radial Basis Function, Least Mean Squares.

Abstract: Chronic Obstructive Pulmonary Disease (COPD) is characterized by airflow limitation and the spirometry is one of the tests that can be used to detect such disease. However there is a great problem related to the different ways of interpreting the values provided by spirometric devices, regarding different guidelines and reference values. Artificial Neural Networks (ANN) can be used to help with tasks of diagnosis as that. This work presents the modeling and analysis of three ANN models to classify subjects with COPD, based on different sets of variables: a set of observed measures from spirometry and a set of interpreted values according to the guideline proposed by the American Thoracic Society. The results shown that it is possible to achieve a good accuracy in the diagnosis of COPD using ANNs, besides these features set conducted the COPD identification problem to a nearly linearly separable classification problem.

1 INTRODUCTION

According to the Global Initiative for Chronic Obstructive Lung Disease (GOLD), Chronic Obstructive Pulmonary Disease (COPD) is a disease state characterized by airflow limitation that is not fully reversible. In (Group, 2004), it is reported that COPD is placed just after cardiovascular and neoplastic diseases as a leading cause of death and morbidity. It also implies a high rate of medical services, as well as hospitalization, frequently, for a long time (Lenfant, 1998).

Diagnosing and monitoring progression of COPD is commonly done by spirometry, considered the gold standard for such tasks. However, differences in the definition of COPD and consensus statements make it difficult diagnosing, as pointed out by (Nathell et al., 2007). In this paper, the authors reported that the use of the European Respiratory Society Guidelines (ERS) has proved to be more effective in detecting patients with COPD than other guidelines, as the GOLD COPD-criteria and the NICE COPD-criteria. The authors also show how the COPD diagnosis is highly dependent on the which guidelines are used for defining the disease.

Artificial Neural Networks (ANN) have been commonly used as classifiers or predictors in several fields of knowledge, including biomedicine. They can be applied in clinical medicine, physiological sign processing and medical image processing. Passold et al. (Passold et al., 1996) resume the benefits of ANN applied in biomedicine in three main points: (*i*) Simulation of human reasoning in diagnosis, based on a given data set; (*ii*) Learning ability in a self-organized way; and (*iii*) High performance when compared to statistical methods. The task of diagnosing diseases, which present some level of difficulty for human consensus, has been successfully solved using ANN (Zhou and Jiang, 2003), (Wadie et al., 2006), (Yan et al., 2006), and (Mehrabi et al., 2009).

In this context, the aim of this article is to study three ANN models to classify subjects with COPD, based on different sets of variables: a set of observed measures from spirometry and a set of interpreted values according to the guideline proposed by the American Thoracic Society, that is commonly included in the software provided by many spyrometers.

The remaining of the paper is structured as follows: Section 2 presents related work, Section 3 describes the ANN modeling. The performance of each model is analyzed in Section 4. Conclusions and future work are presented in Section 5.

de Oliveira Camargo-Brunetto M. and R. Gonçalves A..
Diagnosing Chronic Obstructive Pulmonary Disease with Artificial Neural Networks using Health Expert Guidelines.
DOI: 10.5220/0004234102070214
In *Proceedings of the International Conference on Health Informatics* (HEALTHINF-2013), pages 207-214
ISBN: 978-989-8565-37-2

2 RELATED WORKS

Fontenla-Romero et al. (Fontenla-Romero et al., 2005) presented a new method to classify sleep apnea as: obstructive, central or mixed. However, this classification method requires to analyze a large number of variables during a long time. So, the inputs of the neural network were pre-processed by applying a discrete wavelet transformation on the samples, aiming to reduce and to fix the number of inputs of the classifier.

Basically, the authors used wavelets to extract features from the recorded signals, that become the input of a feed-forward ANN in order to do the classification. Three learning algorithms were used. Conjugate Gradient Method was employed to adjust the weights and the quadratic mean error as the cost function. A similar algorithm, adds a regularization term to avoid hyper-training. Finally, a learning algorithm uses a Bayesian framework and a cross-entropy error function. The latter showed to be the best choice, with an accuracy of 83.78%±1.90%.

Er and Temurtas (Er and Temurtas, 2008) present a comparative study for the realization of the COPD diagnosis using multilayer neural networks. The authors applied two different Multilayer Perceptron (MLP) structures: one with one hidden layer and the other with two hidden layers. The COPD dataset was taken from the Diyarbakir Chest Diseases Hospital from Southeast of Turkey. The dataset contains 155 samples, where 55 are COPD and 100 normal. They analyzed 38 features from laboratory examination. Accuracy of 93.14% was obtained for an MLP ANN with one hidden layer and backpropagation algorithm with momentum constant. Using the Levenberg-Marquardt learning algorithm, the accuracy achieved was 94.46%. Best results were obtained using MLP with two hidden layers, where the accuracies for same learning algorithms were 95.43% and 96.08%.

Mehrabi et al (Mehrabi et al., 2009) used a MLP and a Radial Basis Function Neural Network (RBFNN) to classify patients with COPD and Congestive Heart Failure (CHF). They analyzed 266 patients, being 129 with CHF and 137 with COPD. It was considered 42 clinical variables. A ten-fold cross validation was used to assess the generalization of the classification models and the results obtained were: MLP sensitivity of 83.9% and specificity of 86%, with AUC (Area under the ROC curve) of 0.889 ±0.002; RBFNN sensitivy of 81.8% and specificity of 88.4%, with AUC of 0.924 ±0.01.

Er et al (Er et al., 2010) analyzed several classifiers in order to classify patients among the following chest diseases: COPD, Pneumonia, Asthma, Tuberculosis and Lung Cancer Diseases. The following models were used: MLP (with one and two hidden layers), without and with momentum; Probabilistic Neural Network, Generalized Regression Neural Network, Learning Vector Quantization and RBFNN. A sample of 357 patients was analyzed, where 71 with COPD, 50 with Tuberculosis, 60 with Pneumonia, 44 with asthma, 32 with lung cancer, and 100 normal. It was used 38 variables as input for all ANN models. For the COPD diagnosis, the best results were obtained with MLP with two hidden layers and Probabilistic Neural Network, with accuracy of 88.73%.

The results achieved with the application of ANN as classifiers of diseases have encouraged researchers study and to apply such models to help with tasks of COPD diagnosis. COPD is a chronic disease that requires constant care as well as knowledge of its severity, that is normally done by spirometry. Normally, the devices that perform such exam have softwares that provide different reference values and guidelines to identify the degree of normality of the subject. Since COPD diagnosing is highly dependent on the guideline used for detecting it, ANN could be an alternative to help health care professionals to classify subjects with COPD.

3 ANN MODELING FOR COPD IDENTIFICATION

This research considered three ANN models based on supervised learning: a linear ANN using the LMS algorithm, a Multilayer Perceptron with backpropagation algorithm and a Radial Basis Function Neural Network. A detailed description of each one is presented in Section 3.3. The physiologic measures used for this study were obtained from pulmonary function test and performed at the Physiotherapy Pulmonary Laboratory at State University of Londrina, Brazil.

3.1 Obtaining the Physiological Measures

The pulmonary function test is composed of several exams that provide information about the pulmonary capacity of a subject. It is composed of spirometry, ventilometry, and measures of the inspiratory and expiratory pressures. The spirometry measures the volume, capacity, and pulmonary flow, from respiratory maneuver and compares them with reference values of normality to the evaluated population. For all volumes and capacities measured by the spirometry, there are reference tables, which predict values are based on regression equations. Several guidelines and normality standards have been proposed by scientific societies e.g. ATS (American Thoracic Society), ERS (European Respiratory Society), ALATS (American Latin Thoracic Society), and others. Nowadays, the spirometry has been used in the respiratory physiotherapy as a complementary exam and has been showed very helpful to physiodiagnosis and to plan a therapeutic program. The spirometer used in this work provides a few options of reference values, and the ATS pattern was selected by the healthcare professional that did the exam (Azeredo, 2002).

The obtained values by spirometry provide rich information about the pulmonary function, helping to identify and to qualify the severity of several ventilatory disturbs that are observed by changes in the spirometric values. It must be remembered that the interpretation of the results is highly dependent on the guideline used. The physiological measures used in this study were: Forced Vital Capacity (FVC), Forced Expiratory Volume at the first minute (FEV1), Pick Expiratory Flow (PEF), Forced Expiratory Flow (FEF), Maximum Voluntary Ventilation (VVM), Minimum Inspiratory Pressure (MIP), and Maximal Expiratory Pressure (MEP). Besides those measures, the variables gender, age, and body mass index were also taken into account.

3.2 Identifying Inputs and Output Variables

The physiological measures used were taken from 222 patients, being 80 with COPD (declared by physiotherapists that work in rehabilitation program with them based on clinical exams and spirometry). The clinical and spirometry data derived two datasets for training and testing of different ANN models considered in our experiments. The first dataset (S1) was created with the following variables as input: age, gender, body mass index, FVC(o), FEV1(o), PEF(o), FEF 25-75 %(o), Maximal Ventilatory Volume (MVV(o)), Maximal Inspiratory Pressure (MIP (o)), and Maximal Expiratory Pressure (MEP(o)). The notation (o) means the obtained values by the spirometry. The second dataset (S2) was defined by the quotient between the measured and the expected value for a normal subject (considering its gender and age) following normality equations proposed by ATS, being each input pattern composed of five variables: FVC (%), FEV1(%), PEF(%), FEF25-75%(%), and MVV(%). The input variables were normalized to values between 0 and 1, including the variables age, gender (binary digits 0 and 1) and body mass index. As all ANN models used in this research are based on supervised learning, the output answer of each input is also included in the pattern representation. Two classes were defined: Normal subject and COPD subject, represented by the binary digits 0 and 1, respectively.

3.3 ANN Architectures Analyzed

To devise this research three different ANN architectures were designed to perform the classification task, namely: Linear Neuron with the Least Mean Square (LMS) learning algorithm, Multilayer Perceptron and Radial Basis Function ANN. This section describes the modeling of each architecture and the results describing the performance of the ANNs are presented in the next section.

Methodology for the Learning Process. For the ANNs performance evaluation, a ten-fold cross-validation method was used. Data was divided into ten folds and the training was repeated ten times. Each time, we applied nine folds for ANN training and the remaining fold for validation. The final accuracy was obtained by the average results over the ten validation folds.

For all ANNs, the initial values of the synaptic weights were set to random values between [-0.05,0.05].

To obtain a proper number of epochs for the ANNs training, the following methodology was used: initially it was established an arbitrary upper bound for the number of epochs. Then, for each epoch, it was computed the mean square error (MSE) of the outputs, for the training set as well as for the validation one. Analyzing the MSE curve of validation and the MSE curve of training it was possible to determine if the number of epochs was suitable or should increase. Nevertheless, if during the training an indicative of overtraining is detected (by analyzing the MSE curve of validation), the training is promptly stopped. This approach pointed out that an upper bound for the number of training epochs equal to 100 was sufficient for training the three studied ANNs, considering our datasets.

3.3.1 Linear ANN

This ANN is composed of only one linear neuron, which output is given by the result of the linear combination of synaptic weights and inputs, plus bias. Here, this linear ANN was trained via the Least Mean

Squares (LMS) algorithm developed by Widrow and Hoff (Widrow and Hoff, 1960).

3.3.2 Multilayer Perceptron

The Multilayer Perceptron (MLP) is one of the most studied ANN models. Its importance comes from the addition of hidden layers and the use of a nonlinear activation function (typically a sigmoid function) that allows the ANN deal with problems which classes are non-linearly separated. Both MLPs have one input layer, one hidden layer and one output layer. The number of neurons used at the input layer is the same of the number of variables (10 to the S1 and 5 to the S2) plus one for the bias. MLPs with different numbers of neurons in the hidden layer were tested, varying from 2 to 20, and that one with better performance was selected. Evaluation criteria for the classifiers' performance is presented in Section 3.4.

Training and Testing the MLPs. During the training phase, weights and bias of each MLP are updated according to the scaled conjugate gradient (SCG) method (Møller, 1993). SCG belongs to the class of Conjugate Gradient Methods (CGMs) and has been shown to be considerably faster than standard backpropagation and other CGMs (Møller, 1993). Another important advantage is that SCG is fully automated including no user dependent parameters. The activation function used was the sigmoid logistic.

3.3.3 Radial Basis Function

A Radial Basis Function Neural Network (RBFNN) is composed of an input layer, a hidden layer, and an output layer. Just as in the two previous ANN models, at the input layer, the number of neurons is equal to the number of features. The hidden layer consists of an arbitrary number of RBFs (e.g. Gaussian RBFs), being each one defined by a center position and a dispersion parameter (σ). The output layer is formed by neurons that promote a linear combination of the activations of the hidden layer neurons.

As for the previous ANNs, we designed two Radial Basis Function (RBF) networks for the classification of the sets S1 and S2, with input layer with 10 and 5 neurons, respectively, plus bias.

Training and testing the RBFNN. The RBFNN training process was that one implemented by the MATLAB Neural Network toolbox, which consists of: (*i*) starting with no one neuron, the network is simulated; (*ii*) the input vector with the greatest error is selected; (*iii*) a neuron is added to the hidden layer with weights equal to that input vector and; finally, (*iv*) the output layer weights are redesigned to minimize error. So, an upper bound value for the number of hidden layer neurons must be provided. Looking for a better performance, a grid search procedure was performed, varying the number of hidden neurons from 5 to 30.

The values of the dispersion parameter σ were set in an arbitrary way, receiving changes and combinations at each simulation.

3.4 Evaluating Classifiers' Performance

Four performance metrics were taken into consideration in our analysis, namely: area under curve (AUC), specificity, sensitivity and, accuracy.

AUC is obtained by the integration of the Receiver Operating Characteristics (ROC) curve over a set of thresholds considered. ROC is a technique to visualize, organize and analyze the classifier performance by means of the False positive rate and True positive rate relation, given a set of thresholds. Since AUC is a portion of the area of the unit square, its value will always be between 0 and 1.0. AUC also has important statistical properties and is related to other metrics, such as Gini coefficient (Fawcett, 2006).

Sensitivity and specificity are one approach to quantifying the separation ability of the classifier. While the former is the proportion of true negatives that are correctly identified by the test, the latter describes the proportion of true negatives that are correctly identified by the classifier.

Accuracy, in turn, is defined by the ratio of the number of correct classifications to the number of patterns tested.

Simulation results for MLP and RBFNN were obtained using MATLAB R2010b Neural Network Toolbox, whereas the linear ANN with LMS algorithm was implemented by the authors on the same platform.

4 RESULTS AND DISCUSSION

This section presents the results for simulations of the three ANNs modeled. The results are presented in terms of mean and standard deviation over 30 independent runs. These results were obtained, considering a lower bound of acceptance of 0.5, i.e., if a lower bound of 0.5 is established, outputs from 0.5 on are classified as "1".

4.1 Results of Linear ANN

Tables 1 and 2 present the results for dataset S1 and S2, respectively. The best results are those with highlighted background.

It was observed, from Table 1, that values of learning rate μ outside [0.01;0.7] decrease the performance of the ANN, but inside this interval, the performance keeps stable, achieving the best AUC equal to 0.971 ($\pm$ 0.04), with $\mu = 0.5$.

Table 1: Means and standard deviations of the four metrics evaluated from the results of Linear ANN with dataset S1, using different values for parameter μ.

μ	AUC	Accuracy	Spec.	Sens.
0.01	0.968 ($\pm$0.04)	94.71% ($\pm$4.45%)	89.29% ($\pm$11.27%)	97.78% ($\pm$4.43)
0.1	0.971 ($\pm$0.04)	95.50% ($\pm$4.03%)	91.29% ($\pm$10.26%)	97.87% ($\pm$3.53%)
0.3	0.965 ($\pm$0.04)	95.79% ($\pm$4.08%)	92.08% ($\pm$9.85%)	97.88% ($\pm$3.58%)
0.5	0.971 ($\pm$0.04)	96.03% ($\pm$3.69%)	92.88% ($\pm$8.84%)	97.82% ($\pm$3.71%)
0.7	0.957 ($\pm$0.05)	93.64% ($\pm$5.36%)	88.92% ($\pm$12.65%)	96.29% ($\pm$5.57%)
0.9	0.826 ($\pm$0.16)	73.04% ($\pm$18.15%)	77.08% ($\pm$26.75%)	70.76% ($\pm$32.56%)

Although the results of linear ANN for the dataset S2 were stable for all values for the parameter μ analyzed (Table 2), it is possible to observe a high frequency of decrease in terms of accuracy and sensitivity. Conversely, we can observe an increase in terms of specificity.

Table 2: Means and standard deviations of the four metrics evaluated from the results of Linear ANN with dataset S2, using different values for parameter μ.

μ	AUC	Accuracy	Spec.	Sens.
0.01	0.970 ($\pm$0.03)	94.51% ($\pm$4.41%)	96.00% ($\pm$5.09%)	91.88% ($\pm$9.97%)
0.1	0.965 ($\pm$0.04)	94.19% ($\pm$4.52%)	95.21% ($\pm$5.43%)	92.38% ($\pm$10.21%)
0.3	0.970 ($\pm$0.04)	94.85% ($\pm$4.45%)	95.73% ($\pm$5.60%)	93.29% ($\pm$8.57%)
0.5	0.968 ($\pm$0.03)	94.13% ($\pm$4.55%)	94.62% ($\pm$5.85%)	93.25% ($\pm$8.75%)
0.7	0.969 ($\pm$0.04)	94.57% ($\pm$4.48%)	95.01% ($\pm$5.43%)	93.79% ($\pm$8.64%)
0.9	0.968 ($\pm$0.03)	94.41% ($\pm$4.40%)	94.69% ($\pm$5.85%)	93.92% ($\pm$8.14%)

According to these good results achieved by the linear model, it is possible to infer that with both sets of selected variables, S1 and S2, lead to a nearly linearly separable classification problem.

4.1.1 Results of MLP

Tables 3 and 4 present the results obtained by the Multilayer Perceptron with different number of hidden neurons using datasets S1 and S2, respectively. The best results are those with highlighted background.

As mentioned in Section 3.3, MLPs with different numbers of neurons in the hidden layer were tested, varying from 2 to 20. Nevertheless, we present the results for those that led to the best performances.

Table 3: Means and standard deviations of the four metrics evaluated from the results of MLP with dataset S1, considering different numbers of hidden neurons.

hidden neurons	AUC	Accuracy	Spec.	Sens.
5	0.963 ($\pm$0.04)	95.21% ($\pm$4.12%)	91.42% ($\pm$9.04%)	97.34% ($\pm$4.37%)
6	0.965 ($\pm$0.04)	95.28% ($\pm$3.91%)	91.29% ($\pm$9.90%)	97.53% ($\pm$3.97%)
7	0.963 ($\pm$0.04)	95.16% ($\pm$4.44%)	91.00% ($\pm$10.73%)	97.51% ($\pm$4.00%)
8	0.960 ($\pm$0.05)	94.86% ($\pm$4.51%)	90.12% ($\pm$11.38%)	97.51% ($\pm$3.92%)
9	0.966 ($\pm$0.04)	95.63% ($\pm$3.99%)	92.17% ($\pm$9.21%)	97.58% ($\pm$4.27%)

We can see that MLP also has produced good and stable results for both datasets. While a more complex model, with 9 neurons, presented the best performance for dataset S1, a simpler one, with 3 neurons, achieved the best results for dataset S2.

Table 4: Means and standard deviations of the four metrics evaluated from the results of MLP with dataset S2, considering different numbers of hidden neurons.

hidden neurons	AUC	Accuracy	Spec.	Sens.
3	0.965 ($\pm$0.04)	95.14% ($\pm$4.33%)	95.97% ($\pm$5.22%)	93.67% ($\pm$8.82%)
4	0.962 ($\pm$0.04)	94.57% ($\pm$4.51%)	95.22% ($\pm$5.62%)	93.42% ($\pm$8.58%)
5	0.962 ($\pm$0.04)	94.38% ($\pm$4.82%)	95.32% ($\pm$5.57%)	92.71% ($\pm$10.25%)
7	0.956 ($\pm$0.04)	93.85% ($\pm$4.69%)	94.69% ($\pm$6.10%)	92.33% ($\pm$10.15%)
8	0.963 ($\pm$0.04)	94.92% ($\pm$4.18%)	95.51% ($\pm$5.47%)	93.88% ($\pm$9.06%)

Even presenting similar performance in terms of AUC for both dataset S1 and S2, we can identify decreases in the accuracy and sensitivity in dataset S2, when compared with results obtained with dataset S1. On the other hand, it is possible to notice an increment in terms of specificity.

4.1.2 Results of RBFNN

Tables 5 and 6 present the results of some combinations of values to the parameters σ, using RBFNN with dataset S1 and S2, respectively. As informed in Section 3.3, we performed a grid search procedure to find out a proper number of hidden neurons, varying from 5 to 30. For each configuration, we varied the dispersion parameter σ from 0.1 to 10. An RBFNN with 15 hidden neurons has shown the best results. Due to lack of space, we omitted the results of the other experiments.

Table 5: Means and standard deviations of the four metrics evaluated from the results of RBFNN with dataset S1, using different values for parameter σ.

σ	AUC	Accuracy	Spec.	Sens.
0.1	0.908 (±0.06)	89.05% (±6.04%)	79.71% (±16.04%)	94.32% (±6.49%)
0.5	0.966 (±0.04)	95.95% (±3.77%)	92.58% (±8.87%)	97.84% (±3.69%)
1.0	0.969 (±0.04)	96.21% (±4.02%)	92.25% (±9.61%)	98.44% (±3.10%)
10	0.971 (±0.03)	95.93% (±4.00%)	92.42% (±9.41%)	97.91% (±3.71%)

Regarding the results presented in Table 5, higher values of the dispersion parameter, particularly 1 and 10, presented the best performances in terms of AUC and sensitivity.

Table 6: Means and standard deviations of the four metrics evaluated from the results of RBFNN with dataset S2, using different values for parameter σ.

σ	AUC	Accuracy	Spec.	Sens.
0.1	0.968 (±0.04)	96.24% (±3.63%)	96.11% (±4.95%)	96.46% (±6.74%)
0.5	0.964 (±0.04)	95.61% (±4.07%)	95.21% (±5.75%)	96.33% (±6.63%)
1.0	0.960 (±0.05)	95.16% (±4.27%)	95.03% (±5.78%)	95.38% (±7.30%)
10	0.961 (±0.05)	95.09% (±4.39%)	95.05% (±5.63%)	95.17% (±7.62%)

As already observed in the two previous ANN models, the RBFNN also achieved similar performance in terms of AUC for both dataset S1 and S2, and a slight decrease in accuracy and sensitivity in dataset S2, when compared with results obtained with dataset S1. (Exception for $\sigma = 0.1$). Conversely, we notice an increment in terms of specificity.

4.2 Comparing Models

In this section we present a statistical comparative study of the three ANN architectures analyzed. This study is conducted only looking at the values of AUC and sensitivity. The reason for this is that AUC carries more information about the classifier power, whereas sensitivity is related to false negative rate - *fnr* (type II error), through $sens = (1 - fnr)$, which quantifies cases where a patient is diagnosed as normal when he actually is not, being so, a dangerous diagnosis. Therefore, classifiers with high sensitivity (low *fnr*) are preferable.

We carried out a one-tailed Student t -test with a significance of 0.05 to compare the performance of the classifiers which found the best results (highlighted background in tables) and the results are presented in Table 7.

Table 7: The t-test results regarding Alg.1 – Alg.2 is shown as "+", "–", or "$\sim$" when Alg. 1 is significantly better than, significantly worse than, or statistically equivalent to Alg. 2, respectively.

t-test result	S1		S2	
	AUC	Sens.	AUC	Sens.
LMS – RBF	$\sim$	–	$\sim$	–
RBF – MLP	$\sim$	+	$\sim$	+
MLP – LMS	$\sim$	$\sim$	$\sim$	$\sim$

No significant difference between linear ANN (LMS) and MLP was found for both AUC and sensitivity. It is worth mentioning that MLP is a more complex model and require a more computational intensive learning process than the least mean square algorithm.

RBFNN also did not present significant better performance when compared to MLP and linear ANN in terms of AUC, but produced classifiers with the highest sensitivity. Given these results, we could say that RBFNN seems to be the first option to apply to the COPD identification problem.

4.3 Comparing Feature Sets

We performed a statistical comparative study of the classifiers' performance over the two different datasets, S1 and S2. The former, as described in Section 3.2, consists of normalized raw data of three non-physiological and seven physiological measures. The latter considers five physiological measures that incorporate previous knowledge from health experts. The results are presented in Table 8.

It is possible to see that all classifiers lose the capacity of reduce false negatives when using the set of

Table 8: The *t*-test results regarding Alg. is shown as "+" or "∼" when the performance of the Alg. with S1 is significantly better than or statistically equivalent to the performance when using S2, respectively.

	AUC	Sensitivity
LMS	∼	+
MLP	∼	+
RBF	∼	+

variables S2. One hypothesis is that using predicted values from regression equations as reference values, may add noise to the physiological measure, due to residuals from the regression, which, in turn, can mislead the classifier.

4.4 Identifying Discriminating Variables

We are now interested in finding out what are the most discriminating variables. This information could provide some clues about what physical and physiologic characteristics have more influence in the decision process implemented by the ANNs studied.

As the three analyzed classifiers achieved quite similar performance, we will look at the weights of the linear ANN learned through the least mean square algorithm. The means and standard deviations of the weights obtained by linear ANN for dataset S1 and S2, based on 30 independent runs, are depicted in Figures 1 and 2, respectively.

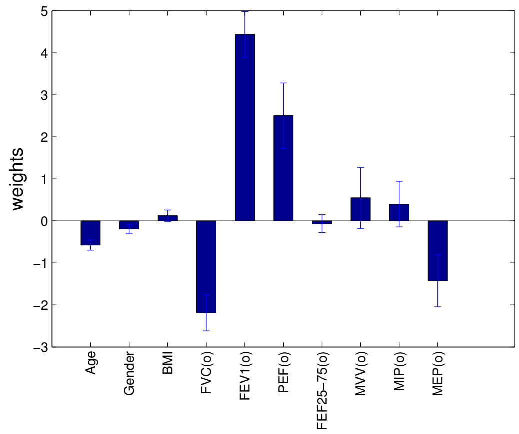

Figure 1: Means and standard deviations of the linear ANN weights for the S1 dataset over 30 executions.

For the dataset S1, it is possible to see that gender and body mass index of a subject have a little influence in the decision process. Additionally, the physiologic measure FEF25-75 (o) are practically ignored by the classifier, although it is worth mentioning its high standard deviation. On the other hand, the most discriminating variables refer to the physiological measures FVC (o), FEV1(o), and PEF(o).

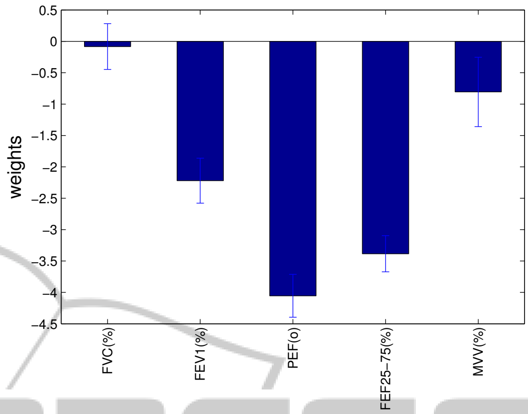

Figure 2: Means and standard deviations of the linear ANN weights for the S2 dataset over 30 executions.

The only neglected variable by the linear ANN in the S2 dataset is the physiologic measure FVC(%), but also has a high standard deviation. That using reference measures, sometimes obtained through regression equation, to guide the physiological measures led to a different classification problem, probably more difficult to deal with, given the results.

5 CONCLUDING REMARKS AND FUTURE WORK

The analysis and classification of physiological measures is not a trivial task. The complexity to develop this kind of research includes accurate data acquisition, selection of a suitable classifier that can provide a good representation of that patterns. This work presented results of modeling, training and testing three different ANN architectures to identify the Chronic Obstructive Pulmonary Disease. Our results highlight the potential of ANNs as a support decision tool for the problem of COPD identification.

The results showed that even simple models as the linear neuron with LMS algorithm had a good performance. It allows us to infer that these sets of selected variables, S1 and S2, conduct the COPD identification problem to a nearly linearly separable classification problem.

Two sets of variables were considered in our analysis: one composed of normalized raw data of three non-physiological (age, gender and body mass index) and seven spirometric measures; and the other consisting of five variables defined as the ratio of the spirometric measures to the expected value for a normal subject following the guideline provided by the American Thoracic Society. It was performed as an attempt to incorporate knowledge from health experts to the dataset. Our results showed that, despite having similar performance in terms of AUC, all classifiers lose the capacity of reduce false negatives when using the second set of variables.

Among the ANNs models analyzed, RBFNNs obtained similar results in terms of classification power, but better performance when looking at the classifiers' sensitivity, for both datasets. This measure tells us that RBFNN classifiers are more likely to avoid false negative diagnosis, i.e., cases when a COPD patient is diagnosed as normal, that may be dangerous.

Such results agree with that obtained by (Mehrabi et al., 2009) and the performance measures obtained in our work are slight better, even considering that other feature set had been considered. The results obtained with the application of ANN in the classification of diseases encourage the study of new applications of such models to help with problems of biomedicine, pointing out the ANN as a powerful technique to help with the understanding and diagnosing diseases.

In this work, ANNs were used only to identify the presence or absence of COPD. As future work, it is intended to apply ANN to classify the level of severity of the disease as well as to support decision on treatment, according to this level.

ACKNOWLEDGEMENTS

This work was supported in part by the National Council for Research and Development of Brazil. We also thank Prof. Dr. Antonio Fernando Brunetto (in Memoriam) by his collaboration as the expert in COPD rehabilitation.

REFERENCES

Azeredo, C. A. C. (2002). *Fisioterapia Respiratória Moderna*. Manole, São Paulo, 4 edition.

Er, O. and Temurtas, F. (2008). A study on chronic obstructive pulmonary disease diagnosis using multilayer neural networks. *Journal of Medical Systems*, 32:429–432.

Er, O., Yumusak, N., and Temurtas, F. (2010). Chest diseases diagnosis using artificial neural networks. *Expert Systems with Applications*, 37(12):7648–7655.

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recogn. Lett.*, 27(8):861–874.

Fontenla-Romero, O., Guijarro-Berdinas, B., Alonso-Betanzos, A., and Moret-Bonillo, V. (2005). A new method for sleep apnea classification using wavelets and feedforward neural networks. *Artificial Intelligence in Medicine*, 34:65–76.

Group, A. W. (2004). Update to the Latin American Thoracic Society (ALAT) Recommendations on Infectious Exacerbation of COPD. *Archivos de Bronconeumologia*, 40:315–325.

Lenfant, C. (1998). Estratégia global para o diagnóstico, a conduta e a prevenção da doença pulmonar obstrutiva crônica. Technical report, GOLD - Global Initiative for Chronic Obstructive Lung Disease.

Mehrabi, S., Maghsoudloo, M., Arabalibeik, H., Noormand, R., and Nozari, Y. (2009). Application of multilayer perceptron and radial basis function neural networks in differentiating between chronic obstructive pulmonary and congestive heart failure diseases. *Expert Systems with Applications*, 36(3, Part 2):6956–6959.

Møller, M. F. (1993). A scaled conjugate gradient algorithm for fast supervised learning. *Neural Networks*, 6(4):525–533.

Nathell, L., Nathell, M., Malmberg, P., and Larsson, K. (2007). COPD diagnosis related to different guidelines and spirometry techniques. *Respiratory Research*, 8.

Passold, F., Garcia Ojeda, R., and Muniz Barreto, J. (1996). Hybrid expert system in anesthesiology for critical patients. In *Electrotechnical Conference, 1996. MELECON '96., 8th Mediterranean*, volume 3, pages 1486–1489.

Wadie, B. S., Badawi, A. M., Abdelwahed, M., and Elemabay, S. M. (2006). Application of artificial neural network in prediction of bladder outlet obstruction: A model based on objective, noninvasive parameters. *Urology*, 68:1211–1214.

Widrow, B. and Hoff, M. E. (1960). Adaptive switching circuits. In *IRE WESCON Convention Record, Part 4*, pages 96–104, New York. IRE.

Yan, H., Jiang, Y., Zheng, J., Peng, C., and Li, Q. (2006). A multilayer perceptron-based medical decision support system for heart disease diagnosis. *Expert Systems with Applications*, 30(2):272–281.

Zhou, Z.-H. and Jiang, Y. (2003). Medical diagnosis with c4.5 rule preceded by artificial neural network ensemble. *IEEE Transactions on Information Technology in Biomedicine, 2003, 7(1): 37-42*, 7(1):37–42.