Proceedings of the 28th IEEE
EMBS Annual International Conference
New York City, USA, Aug 30-Sept 3, 2006

Patients on Weaning Trials from Mechanical Ventilation Classified with Neural Networks and Feature Selection

B. Giraldo, *Member, IEEE*, C. Arizmendi, *Member, IEEE*, E. Romero, R. Alquezar, P. Caminal, S. Benito, D. Ballesteros

Abstract— **One of the challenges in intensive care is the process of weaning from mechanical ventilation. We studied the differences in respiratory pattern variability between patients capable of maintaining spontaneous breathing during weaning trials and patients that fail to maintain spontaneous breathing. In this work, neural networks were applied to study these differences. 64 patients from mechanical ventilation are studied: Group S with 32 patients with Successful trials and Group F with 32 patients that Failed to maintain spontaneous breathing and were reconnected. A performance of 64.56% of well classified patients was obtained using a neural network trained with the whole set of 35 features. After the application of a feature selection procedure (backward selection) 84.56% was obtained using only 8 of the 35 features.**

I. INTRODUCTION

NEURAL networks are very sophisticated modeling techniques capable of modeling extremely complex functions [1, 2]. A simple network has a feed-forward structure: signals flow from inputs, forwards through any hidden units, eventually reaching the output units. Such a structure has stable behavior. Neural networks are applicable in every situation in which a relationship between the predictor variables (independents, inputs) and predicted variables (dependents, outputs) exists, even when that relationship is very complex and not easy to articulate in the usual terms of "correlations" or "differences between groups."

The weaning trial protocols are very important in patients under mechanical ventilation. A failed weaning trial is discomforting for the patient and may induce significant cardiopulmonary distress. When mechanical ventilation is discontinued, up to 25 percent of patients have respiratory distress severe enough to necessitate reinstitution of ventilatory support [3].

The respiratory pattern describes the mechanical function of the pulmonary system and can be characterized by the following time series: inspiratory time (T_I), expiratory time (T_E), breath duration (T_{Tot}), tidal volume (V_T), fractional inspiratory time (T_I/T_{Tot}), mean inspiratory flow (V_T/T_I) and frequency-tidal volume ratio (f/V_T).

The variability of breathing pattern is non random and may be explained either by a central neural mechanism or by instability in the chemical feedback loops [4]. The study of variability in the respiratory pattern has been discussed in [5-8].

The aim of the present work is the analysis with neural networks of the respiratory pattern variability in patients during weaning trials, in order to find differences between patients capable of maintaining spontaneous breathing and patients that fail to maintain spontaneous breathing.

As in many real situations, the suitable variables that describe the problem are partially unknown. When irrelevant variables are present, there may be many different models able to fit the data. But only some of them (those that do not use irrelevant variables) will lead to good generalization performance on unseen examples. However, in general it is not possible to control that irrelevant variables are not used during the training phase to learn the training set. This problem is shared by all modeling techniques, including neural networks. In this situation, feature selection techniques allow to search for a good subset of input features.

II. ANALYZED DATA

Respiratory flow was measured in 64 patients on weaning trials from mechanical ventilation (WEANDB data base). These patients were recorded in the

Manuscript received April 23, 2006. This work was supported in part by *Ministerio de Educación y Ciencia* and *FEDER,* under grants TEC2004-02274, TEC2004-05263-C02-01 and CGL2004-04702-C02-02, and FIS grant 02/0376, from the Spanish Government.

B. Giraldo, C. Arizmendi and P. Caminal are with Dep. of ESAII, Centre for Biomedical Engineering Research, Technical University of Catalonia (UPC), c./ Gargallo, 5, 08028, Barcelona, Spain (e-mail: Beatriz.Giraldo@upc.edu, Carlos.Julio.Arizmendi@upc.edu, Pere.Caminal@upc.edu).
E. Romero, R. Alquezar are with Dep. of LSI, Technical University of Catalonia (UPC), C. Jordi Girona, 1-3, 08034, Barcelona, Spain (e-mail: eromero@lsi.upc.edu, alquezar@lsi.upc.edu).
S. Benito is with Dep. Intensive Care Medicine, Hospital de la Santa Creu I Sant Pau, Barcelona, Spain.
D. Ballesteros is with Hospital Universitario de Getafe, Getafe, Spain.

Departments of Intensive Care Medicine at Santa Creu i Sant Pau Hospital and Getafe Hospital, according to a protocol approved by the local ethic committees. Using clinical criteria based on the T-tube test, the patients were classified into two groups: **Group S (GS)**, 32 patients with **S**uccessful trials after 30 minutes and **Group F (GF)**, 32 patients that **F**ailed to maintain spontaneous breathing and were reconnected after 30 minutes of weaning trials.

Respiratory flow was obtained using a pneumotachograph connected to an endotraqueal tube. The signal was recorded at a sampling frequency of 250 Hz during 30 minutes.

From each recorded signal the aforementioned time series were obtained: inspiratory time (T_I), expiratory time (T_E), breath duration (T_{Tot}), tidal volume (V_T), fractional inspiratory time (T_I/T_{Tot}), mean inspiratory flow (V_T/T_I) and frequency-tidal volume ratio (f/V_T).

III. METHODOLOGY

A. Data Preprocessing

Each one of the seven time series was processed by moving a Running Window (RW), with a width range from 3 to 100 of consecutive breath cycles. Five statistics were calculated for each window: the mean ($\bar{x}$), standard deviation (S), kurtosis (K), skewness (Sk) and interquartile range (IQR) of the value. In this way, 35 new time series were obtained for each patient (Fig. 1), that are the features of the system. The optimal width of the RW was selected in the range from 3 to 100 by using a Mann-Whitney Test. Fig. 2 shows the total p-values for all statistics parameters in all time series and for each window width. The best width was 15 breath cycles, with $p < 0.001$ in all cases. Finally this RW width was used to compute the 35 variables for each patient.

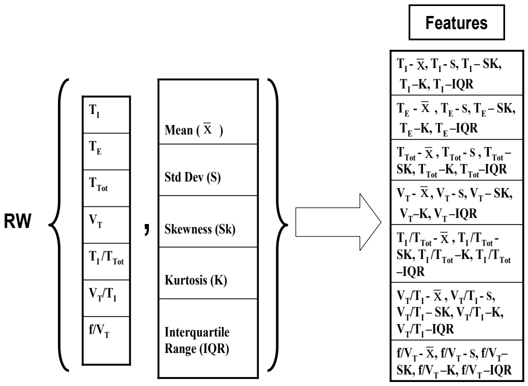

Fig. 1. Representation of 35 features for each patients obtained by Running Window (RW) with seven respiratory time series (T_I, T_E, T_{Tot}, V_T, T_I/T_{Tot}, V_T/T_I and f/V_T and five statistics (mean, standard deviation, skewness, kurtosis and interquartile range) applied.

Next, with the RW width selected, the data of each patient were analyzed independently by applying an algorithm, which automatically determines the best number of clusters for all patients, based on the k-means method. In this study, for all patients, there was a main cluster containing the most part of the patterns (i.e. data points) with a considerable internal cohesion (low intra-cluster variance). Table I shows the mean percentage of data points, represented by the main cluster for Group S and Group F.

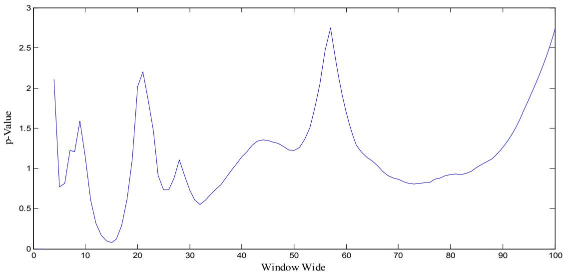

Fig. 2. p-values total for all statistics parameters in all time series and for each window width. The best width was 15 ($p < 0.001$).

TABLE I
PERCENTAGE REPRESENTED BY THE MAIN CLUSTER

	Mean ± *SD* *(%)*
Group S	96,081 ± 3,84
Group F	96,707 ± 3,58

Hence, this result was exploited to perform a dramatic data reduction, by which a single pattern of 35 features was associated with each patient. This pattern was computed as the mean value of the data points in the main (largest) conglomerate of the patient, using the k-means clustering algorithm.

As result of the data preprocessing step, the data set to be classified by the neural network included 64 patterns, one for each patient, 32 from Group S and 32 from Group F.

B. Data Visualization

In order to visualize the data, the 64 patterns were projected onto the plane according to the Sammon mapping [9]. Sammon mapping is a procedure designed to project high dimensional data onto low dimensional spaces preserving the original structure of the data. In particular, every point in the original space is associated to a point in the projected one, so that the distances between the original points are preserved in the projected space as much as possible. Fig. 3 shows the results of the projection in this study. Distances in the original space were preserved at 95%. As it can be seen, there are examples of different classes that are very near between them. This makes the data difficult to classify for a method based on distances

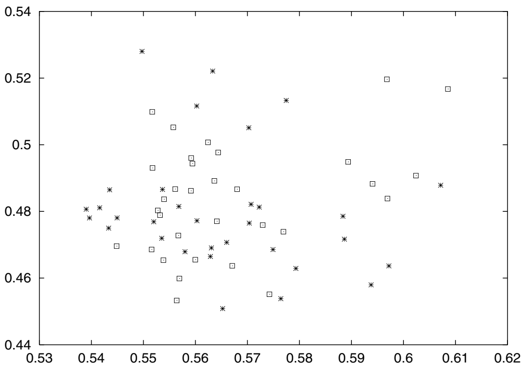

Fig. 3. Projection of the original data onto the plane with the Sammon mapping in patients of Group S and patients of Group F.

like the *k-nearest neighbors* algorithm. However, neural networks are able to learn such a highly non-linear decision surface.

C. *Neural Networks and Features Selection*

Neural networks are used for a wide variety of pattern classification tasks [10]. A pattern is represented by a number of features that form a *d*-dimensional feature vector $\mathbf{x}$ within an input space $X \subseteq R^d$. A classifier therefore realizes a mapping from input space X to a finite set of classes $C = \{1,...,l\}$. A neural network is trained to perform a classification task from a set of training examples $S = \{(\mathbf{x}^{\mu}, t^{\mu}), \mu = 1,...,M\}$ using a supervised learning algorithm. The training set S consists of M feature vectors $\mathbf{x}^{\mu} \in R^d$ each labeled with a class membership $t^{\mu} \in C$. The network typically has as many outputs as classes and the target labels are translated into *l*-dimensional target vectors following a local unary representation. During the training phase the network parameters are adapted to approximate this mapping as accurately as possible (unless some technique, such as early stopping, is applied to avoid over-fitting). In the test phase an unlabeled feature vector $\mathbf{x} \in R^d$ is presented to the trained network and the network outputs provide an estimation of the *a-posteriori* class probabilities for the input $\mathbf{x}$, from which a classification decision is made, usually an assignment to the class with maximum *a-posteriori* probability [1].

In this work, we used a feed-forward 2-layer perceptron architecture (i.e. one hidden layer of neurons and an output layer). For processing the full feature vectors, the networks consisted of 35 inputs, *n* hidden units and 2 output units, where *n* took different values from 5 to 50. Sine and logistic functions were used as activation functions in the hidden layer and the output layer, respectively. The networks were trained with backpropagation. A modified version of the Parallel Distributed Processing (PDP) simulator of Rumelhart and McClelland [11] was employed for the experiments. The weights were initially assigned randomly using a uniform distribution in the interval (-*w*range/2, *w*range/2). Some range values were tried and set independently for the hidden and output layers. Learning rates were adjusted for every architecture and range of weights. The momentum parameter was set to 0, with a maximum number of 1000 training epochs for each run.

For each network tested, 25 independent runs of an 8-fold cross-validation (CV) procedure were carried out, so that many different data partitions were generated, each including 56 patterns in the training set and 8 patterns in the test set. In some cases, 8 patterns not used in the test set were reserved as validation set for early stopping the training phase [1]. In this case, an internal 7-fold CV was performed with the training set of the 8-fold CV, so that 6 folds were used to train, one fold to validate and the remaining fold to test the model.

The network chosen at the end of the training phase was either the one that yielded the best classification result on the validation set among the networks obtained after each training epoch (when a validation set was used) or the one that yielded the best result on the training set (otherwise). Then, the chosen networks were evaluated on the corresponding test sets and the test results were averaged for the different partitions and runs.

First, the experiments with the whole set of 35 attributes were carried out. As previously mentioned, 25 independent runs of an 8-fold CV (without validation set) were performed. After the experiments with the whole set of 35 features, we performed similar cross-validation experiments with different subsets of features, forming groups of 5 statistics related to each one of the seven time series collected. The number of inputs of the networks was set in this case to the dimension of the feature subset.

Finally, in order to determine the best possible subset of input features, feature selection was performed by carrying out several runs of a sequential backward selection method [12, 13]. Features were eliminated one by one, and the network was retrained each time a feature was temporarily removed. Again, 8-fold CV was applied and validation sets were used for early stopping and for selecting the features to remove. Once identified the feature subset providing the best result, 25 runs of an 8-fold CV (without validation set) were performed to obtain the final average test classification result using the selected variables.

IV. RESULTS

Using all the 35 features, the average correct classification rate in the test sets was 64.56%. This result was improved in the case of using some fixed groups of the five statistics obtained from the original time series (Table II), obtaining a best average result of 68.62% using only skewness and standard deviation variables.

However, a much more significant improvement in the

TABLE II

TEST CLASSIFICATION RESULTS FOR SEVERAL GROUPS OF SELECTED VARIABLES USING COMBINATIONS OF FIVE STATISTICS: MEAN, STANDARD DEVIATION, SKEWNESS, KURTOSIS AND INTERQUARTILE RANGE

Feature subsets	***Hidden units***	***Recall***		***Test result ± SD (%)***
		GS	***GF***	
Skewness (*Sk*) and Standard Deviation (*S*)	25	69.75%	67.50%	68.62 ± 1.12
Kurtosis (*K*), Skewness (*Sk*) and Interquartile Range (*IQR*)	25	65.75%	70.75%	68.25 ± 1.21
Standard Deviation (*S*), Skewness (*Sk*) and Interquartile Range (*IQR*)	25	67.75%	67.50%	67.62 ± 1.24
All 35 features	30	64.12%	65.00%	64.56 ± 1.18

TABLE III

TEST CLASSIFICATION RESULTS FOR THE BEST SETS OF SELECTED VARIABLES OBTAINED AFTER SEQUENTIAL BACKWARD FEATURE SELECTION

Feature subsets obtained after sequential backward selection	***Hidden units***	***Recall***		***Test result ± SD (%)***
		GS	***GF***	
$Sk(V_T/T_I)$, $K(V_T)$, $Sk(T_I/T_{Tot})$, $K(T_I)$, $Sk(T_I)$, $IQR(f/V_T)$, $\overline{X}(V_T)$, $IQR(T_{Tot})$	30	76.75%	91.75%	84.25 ± 0.81
$S(V_T/T_I)$, $Sk(f/V_T)$, $S(T_E)$, $\overline{X}(T_I)$, $IQR(f/V_T)$, $Sk(V_T/T_I)$, $IQR(T_I)$, $Sk(T_I)$,$K(T_{Tot})$,$K(V_T)$, $Sk(T_I/T_{Tot})$, $IQR(T_{Tot})$, $\overline{X}(V_T/T_I)$, $\overline{X}(T_I/T_{Tot})$	25	79.12%	83.00%	81.06 ± 1.13
$\overline{X}(f/V_T)$,$IQR(f/V_T)$,$Sk(V_T/T_I)$, $Sk(V_T)$, $IQR(V_T/T_I)$, $IQR(T_I/T_{Tot})$, $\overline{X}(T_I/T_{Tot})$	25	79.75%	80.62%	80.19 ± 0.87

average test result was achieved when applying the sequential backward feature selection process (in which features were not grouped but removed independently). Table III shows the three best average results after backward selection, being 84.25% the best average of well classified patterns in the test set using only 8 of the 35 features.

V. DISCUSSION AND CONCLUSIONS

Neural networks are very sophisticated modeling techniques capable of modeling extremely complex functions and feature selection techniques allow searching a good subset of input features.

Neural networks have been applied in order to analyze the respiratory pattern variability in patients during weaning trials, in order to find differences between patients capable of maintaining spontaneous breathing and patients that fail to maintain spontaneous breathing.

The networks trained with the whole set of 35 features achieved a performance of 64.56%. After the application of feature selection, 84.25% was obtained using only 8 of the 35 features. Although the groups were balanced, a greater facility of learning and generalization was observed in the Group F with 91.75% of successes compared with the Group S with 76.75%.

The optimal width of the RW was computed for the whole set of features, but when the selected features are modified, this optimal RW width may change. The computation of the optimal RW width might be included within the features backward selection process.

It is necessary to make a model with the most important characteristics and to prove the generalization of our system. Additionally, it could be interesting to increase the number of patterns used for training, in order to improve the generalization error, assuring the presence of the subgroup with the most important characteristics.

Given the clustering results, it would be important to carry out a study to obtain the indexes that allowed us to decrease the temporary registry of data.

REFERENCES

[1] C.M. Bishop, *Neural Networks for Pattern Recognition*, Oxford University Press, 1995.

[2] A. Gottschalk, M.C. Hyzer, R.T. Geer, “A comparison of human and machine-based predictions of successful weaning from mechanical ventilation”, *Med.Decis. Making*, 20(2): 160-9, 2000.

[3] M.J. Tobin, “Advances in mechanical ventilation”, *N. Engl. J. Med.,* Vol. 344, N. 26, pp. 1986-1996, 2001.

[4] G. Benchetrit, “Breathing pattern in humans: diversity and individuality”, *Respiration Physiology*, 122, pp. 123-129, 2000.

[5] P. Caminal, L. Domingo, B.F. Giraldo, M. Vallverdú, S. Benito, G. Vázquez, D. Kaplan, “Variability analysis of the respiratory volume based on nonlinear prediction methods”, *Medical & Biological Engineering & Computing*, 42, pp. 86-91, 2004.

[6] M.J. Tobin, M.J. Mador, S.M. Guenter, R.F. Lodato, M.A. Sackner, “Variability of resting respiratory center drive and timing in healthy subjects”, *J. Appl. Physiol*., 65, pp. 309-317, 1988.

[7] E.N. Bruce, “Measures of respiratory pattern variability”, in *Bioengineering approaches to pulmonary physiology and medicine*, Plenum Press, pp. 149–160, 1996.

[8] M.C.Khoo, “Determinants of ventilatory instability and variability”, *Respiratory physiology*, 122, pp. 167–182, 2000.

[9] J. Mao, A.K. Jain, “Artificial neural networks for feature extraction and multivariate data projection”, *IEEE Transactions on Neural Networks*, 6 (2), pp. 296-317, 1995.

[10] E. Fiesler and R. Beale (eds.), *Handbook of Neural Computation*, IOP Publishing Ltd and Oxford University Press, 1997.

[11] D.E. Rumelhart, J.L. McClelland and the PDP Research Group (eds.), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, MIT Press, 1986.

[12] J.M. Sopena, E. Romero, R. Alquezar, “Neural networks with periodic and monotonic activation functions: a comparative study in classification problems”, Proc. Int. Conference on Artificial Neural Networks, ICANN99 , pp. 323-328, 1999.

[13] E. Romero, J.M. Sopena, G. Navarrete, R. Alquezar, “Feature selection forcing overtraining may help to improve performance”, Proc. Int. Joint Conference on Neural Networks, IJCNN-2003, Portland, Oregon, Vol.3, pp.2181-2186, 2003.