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

Support Vector Machine Classification Applied on Weaning Trials Patients

B. Giraldo, *Member, IEEE*, A. Garde, C. Arizmendi, *Member, IEEE*, R. Jané, *Member, IEEE*,
S. Benito, I. Diaz, D. Ballesteros

***Abstract*— One of the most frequent reasons for instituting mechanical ventilation is to decrease patient's work of breathing. Many attempts have been made to increase the effectiveness of the evaluation of the respiratory pattern with the analysis of the respiratory signals. This work proposes a method for the study of the differences in respiratory pattern variability in patients on weaning trials. The proposed method is based on a support vector machine using 35 features extracted from the respiratory flow signal. In this paper, a group of 146 patients with mechanical ventilation were studied: group S of 79 patients with Successful weaning trials and group F of 67 patients that Failed to maintain spontaneous breathing and were reconnected. Applying a feature selection procedure based on the use of the support vector machine with a leave-one-out cross-validation, it was obtained 86.67% of well classified patients on group S and 73.34% on group F, using only 8 of the 35 features. Therefore, support vector machine can be a classification method of the respiratory pattern variability useful in the study of patients on weaning trials.**

I. INTRODUCTION

A Support Vector Machine (SVM) is a binary classifier based on a statistical learning fusion technique [1]. SVM introduced by Vapnik [2] and studied by others [3, 4] is a new and powerful learning methodology that can deal with nonlinear classification a regression. SVM are based on the concept of decision planes that define decision boundaries. A decision plane is one that separates between a set of objects having different class memberships [5-7]. The algorithm can be shown to correspond to a linear method in a high-dimensional feature space non-linearly related to input space [8].

One of the most challenging problems in intensive care is the process of discontinuing mechanical ventilation. Critical-care clinicians must carefully assess the benefits of rapid liberation for mechanical ventilation against the risks of premature trials of spontaneous breathing and extubation. The need for accurate prediction applies to all phases of weaning, beginning with reductions in mechanical support, as patients are increasingly able to support their own breathing, followed by trials of unassisted breathing, which often precede extubation, and ending with extubation [9]. When mechanical ventilation is discontinued, up to 25 percent of patients have respiratory distress severe enough to require reinstitution of ventilation support [10].

The respiratory pattern describes the mechanical function of the pulmonary system and can be characterized by the following features: 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 aim of the present work is the analysis of the respiratory pattern variability in patients during weaning trials, with support vector machine, in order to find differences between patients capable of maintaining spontaneous breathing and patients that fail to maintain spontaneous breathing.

In general, most of the approaches employ methods from the field of signal processing and incorporate the doctor's expertise, in order to reach a satisfactory level of reliability so as to act as a decision support system in respiratory treatments. This research effort aims to exploit the capabilities of classification methods to improve identification and analysis of the respiratory pattern variability.

II. ANALYZED DATA

Respiratory flow was measured in 146 patients on weaning trials with mechanical ventilation (WEANDB data base). These patients were recorded in the Departments of Intensive Care Medicine at Santa Creu i Sant Pau Hospital and Getafe Hospital, according to a protocol approved by

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

B. Giraldo, A. Garde, C. Arizmendi and R. Jané 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,ainaragarde@hotmail.com, Carlos.Julio.Arizmendi@upc.edu, Raimon.Jane@upc.edu).

S. Benito and I. Diaz are 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

1-4244-0033-3/06/$20.00 ©2006 IEEE.

the local ethic committees. Using clinical criteria based on the T-tube test, the patients were classified into two groups: group S, 79 patients with **s**uccessful weaning trials after 30 minutes, and group F, 67 patients that failed 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 pneumotach consists of a Datex – Ohmeda monitor with a Validyne Model MP45-1-871 Variable-Reluctance Transducer. The signal was recorded at a sampling frequency of 250 Hz during 30 minutes.

From each recorded signal the aforementioned several feature 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, consisting of several consecutive breath cycles. Five statistics were calculated for each window: the mean ($\overline{x}$), 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. The optimal width of the RW was selected in the range from 3 to 100 by using a Mann-Whitney test, and the best width was 15, with $p < 0.001$ in all cases.

Next, the data of each patient were analyzed independently by applying a *k*-means clustering algorithm, which automatically determines the number of clusters. In this study, for all patients, there was a main cluster containing the most part of the patterns with a considerable internal cohesion (low intra-cluster variance), corresponding to more than 96%, for each group. The patients were distributed in 80% for training process and 20% for test process.

B. Leave-One-Out Cross-Validation

Leave-one-out cross-validation is an estimator used to assess the performance of learning algorithm (SVM) and can be used to estimate the accuracy of a classifier [11]. This external cross-validation should be undertaken to correct the error committed by choosing randomly the training and testing set. Each sample is removed in turn and the classifier is built using the remaining samples. The class of the removed sample is then predicted using the classifier. This process is repeated for each sample in the original training set with which the classification accuracy is estimated.

The best classification result was obtained with the next 8 features: $S(T_E)$, $\overline{x}\,(T_{Tot})$, $\overline{x}\,(T_I)$, $\overline{x}\,(T_E)$, $S(T_{Tot})$, $IQR(T_{Tot})$, $\overline{x}\,(f/V_T)$, $\overline{x}\,(T_I / T_{Tot})$. These features were used for the final classification process based on support vector machines.

C. Histogram equalization

In order to achieve a reduction of the overlap between successful and unsuccessful patients (group S and group F), the variances of the features might be similar. Unfortunately, these variances cannot always be expected to be similar. In order to solve this problem, an equalization of the histograms of the previously selected features, as a non affine normalization process, is proposed in this paper [12].

Histogram equalization or cumulative distribution function equalization is a general non parametric method to make the cumulative distribution function (CDF) of some given data match to a reference distribution [13].

The principle of this method is to find a non linear transformation to reduce the mismatch of two signals. This transformation maps the distribution of a signal back to the distribution of the reference signal, and it is defined by means of the cumulative distribution functions of the signals in process.

The statistical matching technique applied in the current work matches the cumulative distribution function obtained from one feature and the cumulative distribution function obtained from another feature, both evaluated over the training data.

The designed equalization takes as a reference the feature whereby the minor classification error is obtained on the leave one out cross-validation process which is the $S(T_E)$ feature.

D. Support Vector Machines

Learning based fusion can be treated as a pattern classification problem in which the features obtained with individual classifiers are seen as input patterns to be labeled as group S patients or group F patients obtained from the weaning off mechanical ventilation.

Given a linearly separable two-class training data, the aim is to find an optimal hyperplane that splits input data in two classes: 1 and -1 (the target values that correspond to the group S and group F patients labels respectively) that maximize the distance of the hyperplane to the nearest data of each class**.**

The separating hyperplane for linearly separable data is defined by $x \cdot w_0 + b = 0$, where w_0 is the normal to the hyperplane obtained as a linear combination of a subset of the training data. Data are then classified by computing the sign of $x \cdot w_0 + b = 0$. However, data normally are not separable. In this case, a non-linear decision function is needed, and an extension to non-linear boundaries is achieved by using a specific function called kernel function [3]. The kernel function maps the data of the input space to a higher dimensional space (feature space) by a non-linear transformation. The optimal hyperplane is then constructed in the feature space, creating a non-linear boundary in the input space. The mentioned hyperplane for non-linearly separable data is defined by:

$$f(x) = \sum_{i=1}^{N} \alpha_i t_i \, K(x, x_i) + b \qquad (1)$$

The coefficients α_i and b are determined by solving a large scale quadratic programming problem for which efficient algorithms exist that guarantee global optimum finding. t_i are targets, and $\sum_{i=1}^{N} \alpha_i t_i = 0$. The vectors x_i is the support vectors, which determine the optimal separating hyperplane and correspond to the closest points of each class. N is the number of support vectors and $0 \leq \alpha_i \leq C$, where C is penalty parameter, which allows some flexibility in separating the categories. It controls the trade off between allowing training errors and forcing rigid margins and creates a soft margin that permits some misclassifications. Increasing the value of C, increases the cost of misclassifying points and forces the creation of a more accurate model that may not generalize well. The goal is to find the minimum value of C with which a minimum error classification is obtained.

Furthermore, kernel functions must satisfy some constraints in order to be applicable (Mercer's conditions) [3].

The kernel function used in this work is a radial basis function (RBF) expressed as:

$$K(x_i, x_j) = \exp\left[-\frac{1}{2} \left(\frac{\|x_i - x_j\|}{\sigma} \right)^2 \right] \qquad (2)$$

The best value of the parameter σ was obtained by an internal 10-fold cross-validation applied only over the training set.

IV. RESULTS

The data set was composed of 146 patients (79 patients of group S and 67 patients of group F), of which the 80% (116 patients) formed the training set and the remaining 20% (30 patients) the testing set.

SVM classifier was learned from the training data. Using all the 35 features, the average correct classification rate was 66.67%. Firstly, an internal M-fold cross-validation, shows that the best value of the parameter σ used in the kernel function was $\sigma = 0{,}5$. Secondly, an external leave-one-out cross-validation was applied to get the classification error obtained with each feature, with an SVM trained using one unique feature. This error estimation shown the contribution of each feature to the performance of the classifier, thus it was determined which of the features were the most important and had to be taken into account in the final classification to get the best result.

Once the features were selected, the remaining noisy features were removed, the computational cost reduced and so do the classification error.

Histogram equalization technique was applied to the selected features with the aim of making their accumulation function be similar to the accumulation function of the reference feature that gave the best classification.

The classification percentage obtained after the process developed before was:

Global percentage of well classified patients: 80%.

Percentage of well classified Successful patients: 86.67%.

Percentage of well classified Unsuccessful patient: 73.34%.

A grid-search was done to find the optimum penalty parameter C. As it was checked, the minimum C which gave the best classification error accuracy was selected ($C = 15$), since, thus the cost and the generalization error would be dropped.

Orthogonal projections (Principal Component Analysis PCA) were used as an attempt to visualize the hyperspace data variability on a plane before (Figure 1) and after the classification. Figure 2 and figure 3, show the final

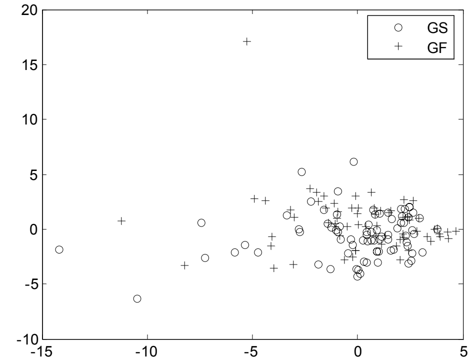

Fig. 1. Successful (GS) and unsuccessful (GF) patients before the classification. Orthogonal projections (Principal Component Analysis PCA) is used to represent hyperspace data on a plane before the classification.

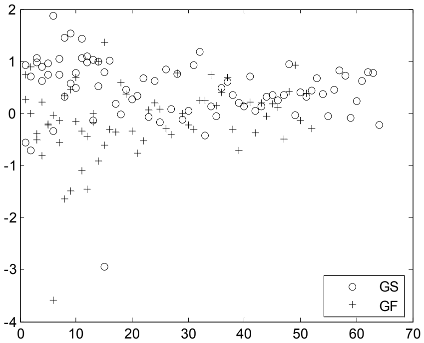

Fig. 2. Training set of successful (GS) and unsuccessful (GF) patients after the classification. Note that appropriate feature selection and histogram equalization process obtained with the training set are previous to the final classification.

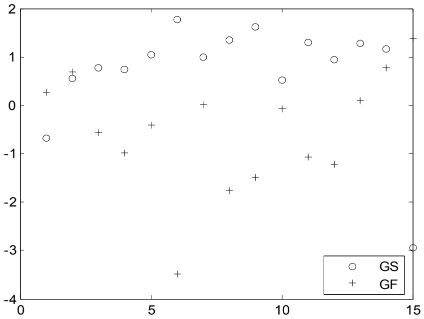

Fig. 2. Test Successful (GS) and unsuccessful (GF) patients after the classification. Note that appropriate variable selection and histogram equalization processes, obtained with the training set are previous to the final classification.

classification of the training set and test set respectively.

V. CONCLUSION

A support vector machine is a robust classification technique capable to classify much overlap data without taking into account the type of the data. Leave one out cross-validation techniques allow searching the best subset of input features and histogram equalization maps the distribution of the selected features back to the distribution of the feature that gives the best classification result.

An innovative method based on support vector machines has 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 SVM trained with the whole set of 35 features achieved a classification error of 66.67%. After the search of the most suitable features and the application of the histogram equalization method taking as reference the $S(T_E)$ feature, 80% was obtained using only 8 of the 35 features. Although the groups are balanced a greater facility of learning and generalization is observed on the part of the Group S with 86.67% of error compared with Group S with 73.34% of error.

These results permit to consider SVM analysis as a promising methodology to study the respiratory pattern variability in patients on weaning trials.

REFERENCES

[1] N. Cristianini, J. Shawe-Taylor, "An Introduction to Support Vector Machines", *Cambridge University Press,* 2000.

[2] V.N. Vapnik, *Statistical Learning Theory.* John Wiley & Sons, New York, 1998.

[3] C.J.C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition, Data Mining and Knowledge Discovery", *Kluwer Academic Publishers, vol. 2, pp. 121-167*, 1998.

[4] K. Veropoulos, N. Cristianini, C. Campbell, "The application of support vector machines to medical decision support: a case study", in *Advanced Course in Artificial Intelligence (ACAI'99)*, Jul. 1999.

[5] B. Crawford, K. Miller, P. Shenoy, R. Rao, "Real-Time Classification of Electromyographic Signals for Robotic Control", *Proceedings of AAI*, pp.523-528, 2005.

[6] T.S. Furey, N. Cristianini, N.Duffy, D.W.Bednarski, M. Shummer , D. Haussler, "Support vector machine classification of cancer tissue samples using microarray expression data", *Bioinformatics* Vo1.6 N. 10, 2000.

[7] G.Georgoulas, C.D. Stylios, P.P. Groumpous, "Predicting the risk of the metabolic acidosis for newborns based on fetal heart rate signal classification usin support vector machines", *IEEE Transactions on biomedical engineering* , Vol.53 , N.5, May 2006.

[8] M. Hearst, "Trends and controversies: support vector machines", *IEEE Intelligent Systems*, Vol. 13 (4), pp. 18-28, 1998.

[9] M. Meade, G. Guyatt, D. Cook, L. Griffith, "Predicting success in weaning from mechanical ventilation", *Chest*, 120, 6, 2001.

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

[11] C. Ambroise, G. J.McLachlan, "Selection bias in gene extraction on the basis of microarray gene-expression data", PNAS Vol. 99, N .10 May 2002.

[12] F. Hilger , H. Ney, "Quantile based histogram equalization for noise robust speech recognition" *Proc. of EUROSPEEC, 2001*, pp. 1135-1138, Aalborg (Dinamarca), September 2001.

[13] R. Balchandran, R Mammone, "Non parametric estimation and correction of non-linear distortion in speech systems", *Proc. IEEE Int. Conf. Acoust. Speech Signal Proc.*, 1998.