POLITECNICO DI TORINO
Repository ISTITUZIONALE

Mixed Neural-Conventional Processing to Differentiate Airway Diseases by Means of Functional Noninvasive Tests

Original
Mixed Neural-Conventional Processing to Differentiate Airway Diseases by Means of Functional Noninvasive Tests / Parvis, Marco; Gulotta, C; Torchio, R.. - In: IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT. - ISSN 0018-9456. - STAMPA. - 50:3(2001), pp. 819-824. [10.1109/19.930460]

Availability:
This version is available at: 11583/1404544 since:

Publisher:
IEEE

Published
DOI:10.1109/19.930460

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

12 October 2024

Mixed Neural-Conventional Processing to Differentiate Airway Diseases by Means of Functional Noninvasive Tests

Marco Parvis, *Senior Member, IEEE*, Carlo Gulotta, and Roberto Torchio

***Abstract*—This paper describes a processing technique that can be used to combine information from different medical analyzes to discriminate between different pathologies that have similar symptoms. The paper is focused on the differentiation between asthma, bronchitis, and emphysema, using only functional noninvasive tests, but the proposed technique can be easily applied to other similar situations where different tests have to be used to identify a pathology. The technique is based on mixed neural-and-conventional processing that not only suggests the pathology, but also estimates the reliability of this suggestion.**

***Index Terms*—Health care, medical, neural networks, processing, uncertainties.**

I. Introduction

THE frequency of lung pathologies is continuously increasing due to several environmental causes, such as air pollution, indoor contaminants, and smoking habits. Such an increase, combined with the increased life expectancy, requires methods to classify the different pulmonary diseases, but which limit the use of costly diagnostic methods as much as possible.

Most functional, noninvasive tests are largely aspecific with respect to pathologies that have similar symptoms at the initial stage, such as asthma, bronchitis and emphysema. For this reason, no test alone allows a reliable discrimination to be performed and, unfortunately, as of yet, no assessed model which is able to meaningfully combine the data exists. This paper describes a mixed neural and conventional approach based on estimation of a "pathology evidence" index on the basis of four lung functional parameters. The proposed approach takes the uncertainty of the single tests into account and flags the processing result with the probability of being the correctly identified pathology.

II. Pathophysiology of Airway Obstruction

The identification of the three airway pathologies [1], on the basis of noninvasive tests, is not an easy task. Airflow tests are usually employed to highlight the presence of emphysema, while respiratory tests, before and after bronchodilating substances, are normally employed to highlight the presence of asthma. Pulmonary emphysema is, in fact, characterized by alveolar destruction and airflow limitation which does not change after the use of bronchodilating substances, but, unfortunately, airflow limitation is also often present in bronchitis patients. Pharmacological reversibility of airway obstruction is a typical feature of asthmatic patients, but chronic obstructive pulmonary disease patients also show variable degrees of response to bronchodilating agents.

A correct diagnosis of asthma, bronchitis, and emphysema can, of course, be reliably obtained by means of clinical, radiological, and functional assessment involving several tests, but this would greatly increase the overall cost and time of the procedure required for the diagnosis.

The aim of this work was, therefore, to verify if a reasonable and accurate prediction could be obtained by combining the results of different simple spirometric data, which are collected before and after pharmacological bronchodilation. Several different tests have been proposed to discriminate between the three pathologies [2]–[4]. After some tests, the authors decided to use four of the tests most commonly found in literature. Two tests concern lung parameters: the residual lung volume (*RV*) and the transfer lung factor for carbon monoxide (*TLCO*). The other two tests are related to the change of two respiratory parameters: forced expired volume in 1s ($\Delta\%FEV_1$) and the specific airway conductance ($\Delta\%sGaw$), before and after inhalation of a broncodilator (200 mg of salbutamol). All the test results were normalized to the standard predicted results according to the European Respiratory Society (ERS) recomendations [5].

III. Data Processing

A. Population

The available data represented a population composed of 158 patients diagnosed according to the American Thoracic Society (ATS) criteria. Of these, 37 were classified as asthmatic, 79 as bronchitic, and 42 as being affected by emphysema. The data were recorded in three different periods with different intrumentation; 96 patients were monitored in 1997 and early 1998, 15 were monitored in January 1999, and the last 47 were monitored in late 1999.

The patients were divided into two groups. One group of 55 patients (13 asthmatic, 29 bronchitic, and 13 affected by emphysema), chosen from among the first 96 patients, was used to estimate the statistical parameters and to train the networks;

Manuscript received May 26, 1999; revised December 19, 2000.

M. Parvis is with the Dipartimento di Elettronica Politecnico di Torino, Torino, Italy.

C. Gulotta and R. Torchio are with Fisiopatologia Respiratoria Az. Osp. S. Luigi, Orbassano, Torino, Italy.

Publisher Item Identifier S 0018-9456(01)04378-9.

0018–9456/01$10.00 © 2001 IEEE

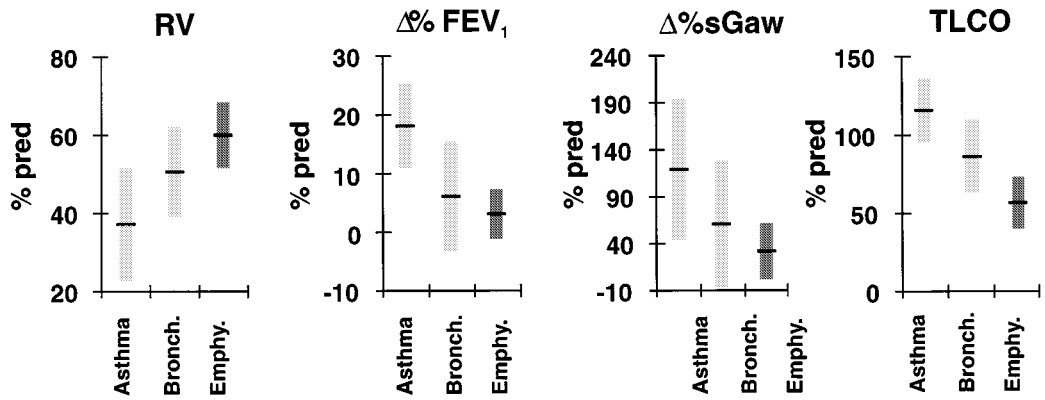


Fig. 1. Mean and standard deviation of the four measured parameters and of the three categories.

data of the remaining 103 patients (24 asthmatic, 50 bronchitic, and 29 affected by emphysema) were used to validate the procedures.

The 55 patients included in the training set were selected by the physicians in order to be representative of the different aspects of the three pathologies. Care was taken to ensure that a reasonable number of examples of the three pathologies were present in the training set to avoid polarized network behavior [6], i.e., to prevent the network from adjusting the weights to describe only the most common pathologies at the expense of the others.

B. Linear Discriminant Score Approach

Fig. 1 shows the mean and standard deviation of the four clinical tests of the patients included in the training set. The results are clustered according to the pathology: a clear correlation between mean values and pathology is visible in each test even though the standard deviations are rather large.

A correlation analysis has shown that the correlations are below 0.5 with the exception of $\Delta\%FEV_1$ and $\Delta\%sGaw$, which reach the value of 0.52.

A patient classification was therefore attempted using a Bayesian approach based on the linear discriminant scores [7]. Such a classification is often used in the medical field (see [8]), and it tries to minimize error probability and cost by assigning a "score" d to each pathology

$$\mathbf{d} = \mathbf{A}\mathbf{x} - \mathbf{c} \tag{1}$$

where

- $\mathbf{x}$ four-row vector, which contains the patient's test results;
- $\mathbf{d}$ three-row vector of the three discriminant scores, one per pathology;
- $\mathbf{A}$ three-row, four-column matrix;
- $\mathbf{c}$ three-row vector.

$\mathbf{A}$ and $\mathbf{c}$ are determined by means of the examples contained in the training set

$$\begin{aligned} \mathbf{A} &= (\overline{\mathbf{x}}_{\mathbf{k}}' S^{-1}) \\ \mathbf{c} &= -\tfrac{1}{2}\overline{\mathbf{x}}_{\mathbf{k}}' S^{-1} x_k + \ln(\mathbf{q}) \end{aligned} \tag{2}$$

where

- $\mathbf{S}$ four-dimensional pooled covariance matrix of the tests of the patients in the training set;
- $\mathbf{x_k}$ four-row three-column matrix, which contains the mean values of each test in the training set, one column per pathology;
- $\mathbf{q}$ three-row vector of the relative frequency of each pathology in the training set, i.e., it is the *a priori* probability of each pathology.

A patient is eventually assigned to the pathology with highest score. The score-based classification was able to guess the right diagnosis in 43 cases (78%) within the training set. Although such a value is reasonably high for this kind of diagnosis, the number of errors (12 cases corresponding to about 22%) is too high for the method being currently used. Statistical procedures are available to validate each result so that doubtful diagnoses could be discovered, but a simpler approach would be desirable.

C. Neural Processing

No commonly accepted analytical model exists that can be used to combine the test results to obtain the diagnosis. However the available examples can be used to train a multi layer perceptron (MLP) neural network, thus overcoming the lack of an analytical model and the problems related to unknown population distributions.

Several network structures can be adopted; all the networks have to have four inputs (the four clinical values) but one can decide to have either one single, three-level output (i.e., a single output which assumes three values that correspond to the three pathologies), or three outputs that activate each one in the presence of a specific pathology.

The authors decided to employ the latter solution, which has two main advantages: a) the three separated outputs allow an easier result interpretation in the presence of doubtful cases to be obtained, and b) the network can be designed in the form of three completely separated sub-networks, which are easier to train.

Two neurons in the hidden layer were found to be sufficient for the networks that had to recognize asthma and emphysema, while three neurons were required for the network which had to recognize bronchitis. All the networks were trained

by employing a gradient-descend, back-propagation (Levenberg–Marquardt) algorithm [9].

Each network was trained to produce an unary output for patients who present the pathology the network had to recognize, and a zero output otherwise. This kind of binary training, where the examples presented to the network have either zero or one target, tends to produce a switching network, especially when the training set is limited [1]. This switching behavior is not correct in our situation, where there are examples of patients who have clinical parameters on the borderline. Such an ambiguity should be reflected also in the values of the MLP output.

This problem can be addressed either by adding noise to the weight values during the training [11] or by adding noise to the network inputs [12]. The authors employed a solution similar to the one described in [13] using an algorithm that takes the uncertainty presence into account [1]. The algorithm acts by replacing each element of the original training set with a sequence of similar new elements that are composed of values which are different from the original ones, within the expected uncertainty of each input parameter.

The three outputs were eventually sent to a competitive layer (CL), which simply employs a winner-takes-all strategy, i.e., it selects the network with the highest output, regardless of its actual value.

By employing the three MLPs and the CL, the network gives 47 correct results (85%) within the training set with only eight errors (15%). The combination of MLP and CL therefore performs better than the algorithm based on the discriminant scores, but still suffers from a high number of errors. Such a high number of wrong diagnoses is due to the nature of the CL. The CL always produces a winner even though all the competitors have a very low value, i.e., even though none of the MLPs have actually recognized their pathology. This behavior can be avoided by employing a modified CL, or some form of more complex algorithms, that is capable of highlighting the presence of doubtful winners.

D. Second-Level Conventional Processing

1) Output Validation Using the Guard Neuron: The CL behavior, which always produces a winner regardless of the winner value, can be modified by adding a guard input (or guard neuron) to the CL inputs. This guard input is a fixed input, set at a suitable level, which wins when all the other inputs are lower than its guard level. The guard level selection should be performed by trying to balance between the number of errors which can be avoided and the number of good results which are missed due to the activation of the guard.

Fig. 2 shows the guard neuron effect as a function of its value for the patients within the control set. The three lines represent the number of erroneous diagnoses, the number of correct diagnoses and the number of "unreliable" diagnoses that trigger the guard neuron. As expected, as the guard level increases, the number of errors decreases. A reduction of the erroneous diagnoses to 2% can be obtained by employing guard levels above 0.8, but at the expense of about 60% of unclassified patients.

2) Output Validation Using the "Evidence Indexes": The guard neuron allows one to recognize conditions where none

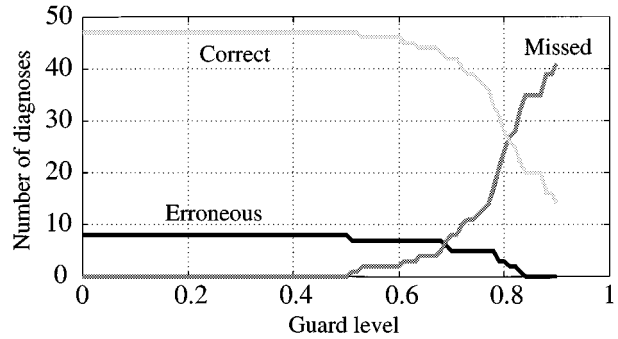


Fig. 2. Erroneous diagnoses (thick line) and unclassified patients (thin line) as a function of the guard level.

of the networks activate, but does not highlight situations where more networks activate with similar values. In addition, the guard neuron approach cannot highlight borderline cases where the uncertainty, which affects the clinical test values, can greatly alter the MLP outputs. Such problems can be reduced by replacing the CL with an algorithm which highlights the evidence of one pathology with respect to the others and takes the uncertainty presence into account. A simple possibility is to compute three evidence indexes by multiplying the output of each network by the complements of the other two

$$e_k = n_k \prod_{j \neq k} (1 - n_j)$$
$$k, j = \langle asthma, bronchitis, emphysema \rangle \tag{3}$$

where k is a pathology index which uses a modulo-three algebra (i.e, if $k = emphysema$ then $k + 1 = asthma$).

Each evidence index can be tagged with its reliability by computing its expected uncertainty as a function of the actual input uncertainties. The uncertainty of the network inputs, i.e., the uncertainties of the clinical parameters, can be estimated according to the ATS criteria. The authors employed an uncertainty of 3% of the expected range of each parameter. Such a value should take all the uncertainty contributions into account and is rather larger than the observed inter and intra-operator variability on a single patient.

The sensitivities of each output with respect to each input depend on the patient's parameter combination and can be numerically computed by examining the network outputs in the presence of small changes of each of the input parameters. A linear approximation of network behaviors can be used since the input uncertainties are small and the network functions do not contain discontinuities. Therefore, the output standard uncertainty can eventually be computed according to the conventional uncertainty propagation rules [14]

$$u_c^2(e_i) = \sum_j s_{ij}^2 u^2(p_j) \tag{4}$$

where

$u_c(e_i)$ combined standard uncertainty of ith evidence index, i.e., the expected standard deviation of ith evidence index;

$u(p_j)$ standard uncertainty of the jth clinical test;

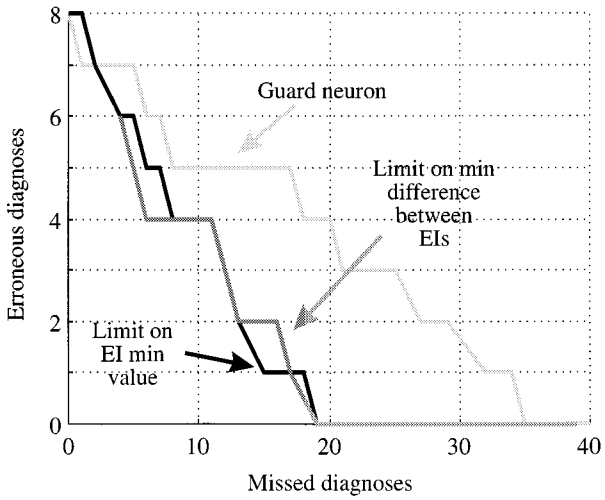


Fig. 3. Good diagnoses versus errors for different criteria.

s_{ij} sensitivity coefficient of the ith evidence index with respect to the jth clinical test.

The covariance is neglected, since the four clinical parameters are measured using independent procedures and devices, and, thus, their uncertainties can be considered statistically independent.

The three evidence indexes condense the amount of information contained in the four original clinical parameters in a structured form, which is easier to manage with respect to the raw data and allows the physician to comfortably carry out the diagnosis.

In addition, the uncertainty associated with each index allows one not only to give an on/off response, but also to guess the reliability of the pathology detection.

The physician can guess the reliability of the diagnosis by using at least three different criteria

1) the actual value of the highest evidence index (low values correspond to less reliable diagnoses);
2) the actual difference between the two higher indexes (low differences correspond to pathologies with similar probabilities of being recognized);
3) the uncertainty associated to each index (that can be used to discriminate between indexes with similar values, but which originate from more or less significant combinations).

Each of these criteria can be used to employ a validation by means of a threshold with an effect which is similar to the guard neuron. Fig. 3 shows the number of good diagnoses versus the number of errors for two criteria that combine index values and uncertainties: the highest evidence index minus its uncertainty, and the difference between the two higher indexes minus the sum of their uncertainties. For comparison, the figure also shows the trace obtained with the guard neuron.

The methods based on the evidence index exhibit similar results and perform better than the method based on the guard neuron. This figure allows one to select the desired compromise between the number of accepted errors and the number of missed diagnoses.

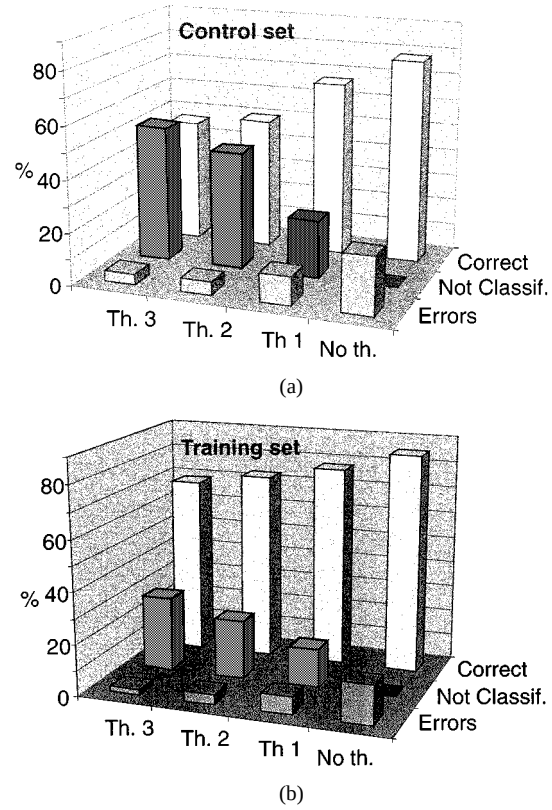


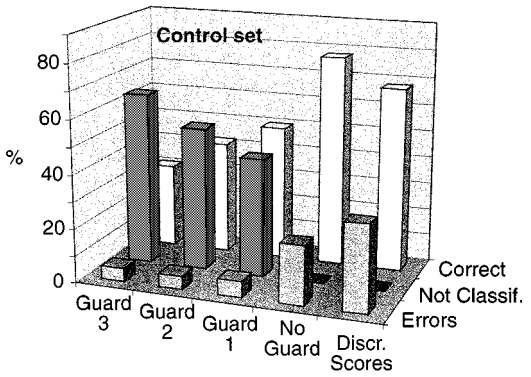
Fig. 4. Evidence index performance on control (left) and training (right) sets with three thresholds and in the absence of thresholds.

The two criteria can of course be mixed to find the combination that seems to be the most suitable to reduce the errors while still maintaining the number of missed diagnoses at low value. After some tests, the authors employed two contemporaneous criteria, discarding results with either the evidence index minus uncertainty below a predefined threshold, or the difference between indexes minus their uncertainties below a second threshold. Three couples of thresholds were selected which corresponded to 4 (7%), 2 (4%), and 1 (2%) erroneous diagnoses within the training set.

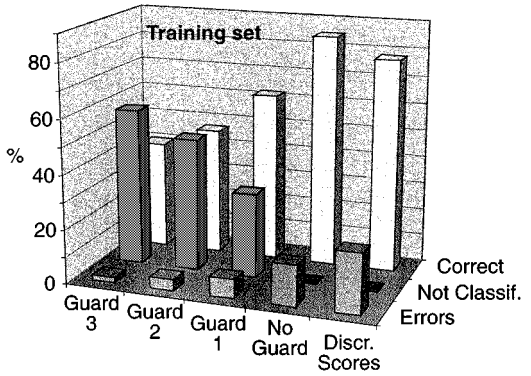
IV. Experimental Results

Fig. 4 shows the performance of the proposed algorithm for the three threshold choices, plus the results obtained in the absence of thresholds. In the absence of thresholds, the system gives 23 (22%) errors and 80 (78%) correct results. Depending on the threshold selection, it is possible to reduce the number of erroneous diagnoses within the control set to 4% at the expense of about 50% unclassified patients, or have 11% errors with 22% unclassified patients. The figure also shows, on the right, the results obtained within the training set with the same thresholds. Within the training set, the results are obviously better than in the control set, especially on the number of unclassified patients, thus suggesting that if more examples were available for both training and testing, an enlargement of the training set could be useful to improve the overall behavior.

The results obtained by the neural network plus the guard neuron, and by the algorithm based on the discriminant scores,



(a)



(b)

Fig. 5. Neural network and discriminant score performance on the control (left) and training (right) sets.

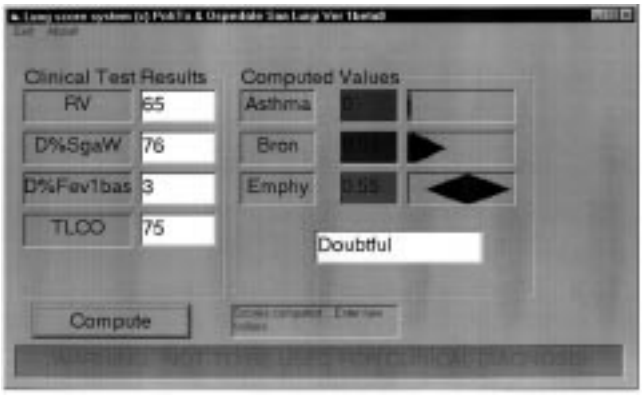


Fig. 6. Panel of the program which computes the evidence indexes and their uncertainties.

are shown in Fig. 5 for comparison purposes. The evidence index method performs better than the other solutions, both in the training set and the control set. A comparison for an equal number of erroneous diagnoses shows that the number of not classified patients is 10-15% higher with the guard neuron than with the evidence index.

The score-based method, which can only be compared with the neural network in the absence of the guard neuron, produces nearly 50% more erroneous diagnoses than the neural network both in the training set (22%) and in the control set (32%).

A simple program has been designed to allow an easy determination of the evidence index. The program, whose graphical interface is shown in Fig. 6, has been coded in VisualBasic™ and is designed to be used in Windows9x/NT™ environments. The program takes the four values that correspond to the four clinical parameters and computes the three evidence indexes with the estimated uncertainties. The network weights and parameter uncertainties are obtained from a file which can be updated by the program that is used for the network training.

V. Conclusion

The discrimination between airway diseases with similar symptoms at early stages can be reliably obtained using a complete clinical, radiological, and functional assessment. However, such a discrimination is much more difficult to obtain when only functional, noninvasive tests have to be employed to avoid unnecessary stress for the patients and to reduce the time required for diagnosis.

This paper has presented a possible procedure to obtain such a discrimination which is based on four simple respiratory tests. The four test results are sent to three MLPs trained to recognize the three pathologies. The network outputs are then combined to define the diagnosis. Two different methods have been presented. The most interesting results are obtained with the method which estimates the evidence index of each pathology and its uncertainty. Starting from these values, each patient is tagged as either not classifiable or affected by one of the three pathologies. The classification is performed by employing a set of thresholds that can be chosen either to reduce the number of erroneous diagnoses, at the expense of a greater number of unclassified patients, or to reduce the number of unclassified patients at the expense of a greater number of erroneous diagnoses.

The proposed algorithm has been trained on a population of 55 patients and tested on another population of 103 patients. Depending on the threshold choice, an error rate in the range of 4% to 10% has been obtained in the control set with a rate of unclassified patients in the range of 50% to 22%. A simple program has been developed which implements the algorithm and can be used to quickly estimate the patient's situation and decide if other tests should be performed.

An analysis of the performance difference within the training and control sets suggests that even better results could be obtained by enlarging the training set to better represent the different kinds of pathologies. The authors are collecting new data to verify this possibility and will update the results as soon as a reasonable number of new examples becomes available.

References

[1] R. Crystal and J. B. West, Eds., *The Lung Scientific Foundations.* New York: Raven, 1991, vol. 2.

[2] N. Meslier, J. Racineaus, and P. Six, "Diagnostic value of reversibility of airway obstruction to separate asthma from chronic bronchitis," *Eur. Respir. J.*, no. 2, pp. 497–505, 1989.

[3] H. Magnussen, "Single breath diffusion capacity for CO in bronchial asthma and emphysema," *Prog. Respir. Res.*, no. 16, pp. 147–151, 1981.

[4] American Thoracic Society, "Standards for the diagnosis and care for patients with chronic obstructive pulmonary disease and asthma," *Amer. Rev. Respir. Dis.*, no. 136, pp. 225–244, 1987.

[5] "Standardised lung function testing," *Eur. Resp. J.*, no. 6, suppl. 16, pp. 1–86, 1993.

[6] R. Anand, K. G. Mehrotra, C. K. Mohan, and S. Ranka, "An improved algorithm of neural network classification of imbalanced training sets," *IEEE Trans. Neural Networks*, vol. 4, pp. 962–969, Nov. 1993.

[7] A. A. Afifi and S. P. Azen, *Statistical Analysis: A Computer Oriented Approach*. San Diego, CA: Academic, 1979.
[8] A. Blinowska, G. Chatellier, J. Bernier, and M. Lavril, "Bayesian statistic as applied to hypertension diagnosis," *IEEE Trans. Biomed. Eng.*, vol. 38, pp. 699–706, July 1991.
[9] M. T. Hagan and M. B. Menhaj, "Training feedforward networks with the Marquadt algorithm," *IEEE Trans. Neural Networks*, vol. 5, pp. 989–993, Nov. 1994.
[10] M. Parvis, "Training and use of neural networks in the presence of uncertainties," in *Proc. IEEE Workshop Emerg. Techn., Intell. Meas. Virtual Syst. Instrum. Meas.*, St. Paul, MN.
[11] A. F. Murray and P. J. Edwards, "Enhanced MLP performance and fault tolerance resulting from synaptic weight noise during training," *IEEE Trans. Neural Networks*, vol. 5, pp. 792–802, Sept. 1994.
[12] L. Holström and P. Koistinen, "Using additive noise in back-propagation training," *IEEE Trans. Neural Networks*, vol. 3, pp. 24–38, Jan. 1992.
[13] C. Gulotta, M. Parvis, and R. Torchio, "Evaluation of surgical risks by means of neural networks in the presence of uncertainties," *Meas.*, no. 23, pp. 171–178, 1998.
[14] *ISO Guide to the Expression of Uncertainty in Measurements*, 1995.

Marco Parvis (M'98–SM'01) was born in Italy in 1958. He received the M.S. degree in electrical engineering in 1982 and the Ph.D. degree in Metrology in 1987, both from the Politecnico di Torino, Italy.

He is now a professor of electronic measurements at the Politecnico di Torino. His main fields of interest are intelligent instrumentation, signal processing, and biomedical measurements. He is now working on new sensors for mechanical and chemical quantities and on the development of distributed measuring systems.

Carlo Gulotta was born in Italy in 1951. He received the M.S. degree in medicine in 1979, specialized in Pneumonology in 1982, and in Internal Medicine in 1991, all from the Turin University, Italy.

He is now Head of Respiratory Pathophysiology and Chief of Emergency Department at the Ospedale S.Luigi, Orbassano Turin, Italy. His main fields of interest are pulmonary gas exchange and respiratory care. He is now working at applications of these fields to critical care medicine.

Roberto Torchio was born in Italy in 1955. He received the M.S. degree in medicine in 1980 and specialized in Respiratory Pathophysiology in 1984, both at Turin University, Italy.

He is now Head of the Cardiopulmonary Functional Diagnostic Laboratory at the Ospedale S. Luigi, Orbassano, Turin. His main fields of interest are lung function and cardiopulmonary tests. He is now working at routine applications of lung function tests with regard to respiratory mechanics and pulmonary gas exchange.