COMPUTER METHODS AND PROGRAMS IN BIOMEDICINE 105 (2012) 183–193

journal homepage: www.intl.elsevierhealth.com/journals/cmpb

Machine learning algorithms and forced oscillation measurements applied to the automatic identification of chronic obstructive pulmonary disease

Jorge L.M. Amaral [a], *Agnaldo J. Lopes* [b], *José M. Jansen* [b], *Alvaro C.D. Faria* [c], *Pedro L. Melo* [c,*]

[a] *Department of Electronics and Telecommunications Engineering, State University of Rio de Janeiro, Rio de Janeiro, Brazil*
[b] *Pulmonary Function Laboratory, Pedro Ernesto University Hospital, State University of Rio de Janeiro, Rio de Janeiro, Brazil*
[c] *Biomedical Instrumentation Laboratory, Institute of Biology Roberto Alcantara Gomes and Laboratory of Clinical and Experimental Research in Vascular Biology (BioVasc), State University of Rio de Janeiro, Rio de Janeiro, Brazil*

ARTICLE INFO

Article history:
Received 1 April 2011
Received in revised form
15 August 2011
Accepted 22 September 2011

Keywords:
Clinical decision support
Artificial intelligence
Classification
Forced oscillation technique
Respiratory system
Chronic obstructive pulmonary disease

ABSTRACT

The purpose of this study is to develop a clinical decision support system based on machine learning (ML) algorithms to help the diagnostic of chronic obstructive pulmonary disease (COPD) using forced oscillation (FO) measurements. To this end, the performances of classification algorithms based on Linear Bayes Normal Classifier, *K* nearest neighbor (KNN), decision trees, artificial neural networks (ANN) and support vector machines (SVM) were compared in order to the search for the best classifier. Four feature selection methods were also used in order to identify a reduced set of the most relevant parameters. The available dataset consists of 7 possible input features (FO parameters) of 150 measurements made in 50 volunteers (COPD, $n = 25$; healthy, $n = 25$). The performance of the classifiers and reduced data sets were evaluated by the determination of sensitivity (Se), specificity (Sp) and area under the ROC curve (AUC). Among the studied classifiers, KNN, SVM and ANN classifiers were the most adequate, reaching values that allow a very accurate clinical diagnosis (Se > 87%, Sp > 94%, and AUC > 0.95). The use of the analysis of correlation as a ranking index of the FOT parameters, allowed us to simplify the analysis of the FOT parameters, while still maintaining a high degree of accuracy. In conclusion, the results of this study indicate that the proposed classifiers may contribute to easy the diagnostic of COPD by using forced oscillation measurements.

1. Introduction

Chronic obstructive pulmonary disease (COPD) is a major cause of chronic morbidity and mortality throughout the world [1]. According to WHO estimates, 80 million people have moderate to severe COPD. More than 3 million people died of COPD in 2005, which corresponds to 5% of all deaths globally [2]. The chronic airflow limitation characteristic of COPD is caused by a mixture of small airway disease (obstructive bronchiolitis) and parenchymal destruction (emphysema) [1]. There is an agreement in the literature that new measurement technologies that are able to detect COPD in early stages would contribute to decreasing medical and economic burdens [3].

* *Corresponding author.*
E-mail addresses: plopes@uerj.br, plopeslib@gmail.com (P.L. Melo).
0169-2607/$ – see front matter
doi:10.1016/j.cmpb.2011.09.009

Submitting a physical system to forced oscillations is a very general approach to the investigation of its structure and/or properties [4]. Its application to respiratory mechanics was first proposed by DuBois et al. [5]. This method, known as forced oscillation technique (FOT), consists of applying small sinusoidal pressure variations to stimulate the respiratory system at frequencies higher than the normal breathing frequency and measuring the flow response. This method characterizes the respiratory impedance and its two components, respiratory system resistance (R_{rs}) and reactance (X_{rs}). The method is simple and requires only passive co-operation and no forced expiratory maneuvers. Recently, this technique has been successfully applied in the detection of early respiratory changes in smokers [6].

Although obtaining respiratory impedance values is easy, the resulting values are difficult to understand by clinicians as they are based on an electrical equivalent circuit model of the respiratory system. In the context of a diagnosis framework, the interpretation of resistance and reactance curves, as well as the derived parameters measured by the FOT, requires training and experience, and is difficult task for the untrained pulmonologist.

Methods based on machine learning (ML) have been widely used to develop classifiers. These systems can extract information from different classes of signals after having been trained to perform this specific task by learning from examples. In respiratory mechanics, ML proved to be useful as a pattern recognition method to optimize alarms of anesthesia breathing circuits [7], detection of upper airway obstruction [8], esophageal intubation [9], assessment of lung injury [10], static compliance in animal models [11] and the evaluation of spirometric exams [12]. Recently, a severity classification for idiopathic pulmonary fibrosis by using fuzzy logic was proposed [13].

2. Background

Previous works [14,15] have compared groups of controls and COPD patients observing clear modifications in FOT parameters. However, categorization of pulmonary diseases by looking at the plotted curves of respiratory impedance or derived parameters can prove a difficult task for the untrained pulmonologist. This raises the question: an ML based approach to the analysis of FOT data can provide an efficient method to recognize COPD? In fact, only two recent conference papers have addressed this question [16,17].

In the work of Baruá et al. [16], an artificial neural network (ANN) was used to recognize and classify the diseases of the central and peripheral airways. The authors used IOS measurements and a feedforward ANN that was trained by the backpropagation algorithm. After supervised training, the classifier produced a 98.47% and 61.53% correct classification rate when the same data and a new set of unseen data were used, respectively. It was pointed out that the proposed classifier could be further improved with the inclusion of more training samples combined with fuzzy logic decision rules.

In a latter work of the same group [17], a classifier based on ANN was capable of distinguishing between relatively constricted and nonconstricted airway conditions in asthmatic children. The performance of the classifier was evaluated by two methods: (1) using all of the patterns during training as well as in the feed-forward stage and (2) using only 60% of the data set during training and with the remaining 40% as unseen patterns. The classification accuracies obtained were 95.01% and 98.61%, respectively. The authors concluded that ANNs can successfully be trained with the impulse oscillation system (IOS) data, enabling them to generalize the IOS parameter relationships to classify previously unseen pulmonary patterns. The two cited studies used an IOS, which has differences from the classical FOT, including data processing and the parameters used to interpret raw data [18,19]. In addition, from a system identification point of view, the impulse excitation signal used in IOS is a much worse excitation signal than a Multisine used in FOT. This difference is associated with a worse crest factor in the impulse signal.

In this context, we observed that there was no data in the literature concerning the use of ML algorithms associated with classical FOT measurements to aid clinicians in the identification of COPD. To contribute to elucidate this question, our group recently investigated this possibility using the classical FOT associated with a classifier based on ANN [20]. Two feature selection methods (the analysis of the linear correlation and forward search) were used in order to identify a reduced set of the most relevant parameters. Two different training strategies for the ANNs were used and the performance of resulting networks were evaluated by the determination of accuracy, sensitivity (Se), specificity (Sp) and AUC. The ANN classifiers presented high accuracy ($Se > 0.9$, $Sp > 0.9$ and $AUC > 0.9$) both in the complete and the reduce sets of FOT parameters. This indicates that ANNs classifiers may contribute to easy the diagnostic of COPD using FOT measurements. Although these results were very promising, this initial work was limited to the investigation of an ANN based classifier because we were interested in a direct comparison with the two previously cited works.

The purpose of the present study is to evaluate the performance of several ML algorithms in the development of an automatic classifier to help the diagnostic of COPD using forced oscillation measurements.

The paper is organized as follows: a discussion of the design principles and implementation goals is presented in the next section. The healthy group and the COPD group are characterized in Section 4, along with a description of the measurement protocol. This section also presents the evaluated classifiers and describes the methods used for performance evaluation, comparisons among classifiers and feature selection. Section 5 presents the results and Section 6 discusses the results with respect to the search for the best classifier and parameters. Section 7 summarizes the main outcomes of this investigation and points to future steps in this research topic.

3. Design considerations

3.1. *Classification system*

The basic structure of a classification system is the input, the classifier and the output. In the present work, the inputs are the parameters provided by the FOT, the classifier is one of the