



A Decision Support System for Chronic Obstructive Pulmonary Disease (COPD)

Ernesto Iadanza and Vlad Antoniu Mudura

Abstract

Obstructive chronic obstructive pulmonary disease (COPD) is a respiratory disease characterized by a chronic air flow limitation and associated with major economic and social problems. In an attempt to find a solution to these problems, numerous systems of clinical decision support for the management of patients with COPD have been developed in recent years. In particular, systems based on machine learning algorithms have been developed with the aim of monitoring the health status of patients and foreseeing and preventing exacerbations and hospital admissions. An in-depth research into scientific literature has shown that, in the state of the art, these goals have not yet been met and the performance of the current systems is not clinically acceptable. The aim of this work is the design and implementation of a new clinical decision support system that can at least partially fill the gaps present. The first step in the work was to try to replicate the performance of support systems for similar decisions, already present in scientific literature. Using the physiological parameters acquired by 414 patients using respiratory function tests, two predictive models were made using the same machine learning algorithms (neural network and support vector machine). The performance obtained was comparable to those of the scientific literature. The next step was to create a new predictive model, with superior performance to previous models. The machine learning algorithm chosen is C5.0. The performance obtained was significantly better than the two previous models. The new predictive model was implemented within a user interface, implemented in Java

programming language, the COPD Management Tool. The software developed allows the evaluation and classification of the results of respiratory performance tests, with excellent performance, compared to the current state of the art and can therefore be used in many clinical applications.

Keywords

DSS • COPD • Machine learning

1 Introduction

Chronic Obstructive Pulmonary Disease (COPD) is a respiratory disease characterized by a chronic airflow limitation and associated with major economic and social problems. COPD is classified as the fourth leading cause of death in the world and in absence of countermeasures aimed to reduce risk factors it is expected to become the third leading cause of death by 2030 [1]. In 2015 there were 3.2 million deaths associated with COPD and the estimated global prevalence of COPD was about 175 million [2]. From an economic point of view, only in the United States, the annual estimated costs associated with COPD are about 50 billion dollars. In the next years, costs are expected to rise dramatically together with prevalence. Costs increase with increasing severity of the disease and most of them are linked with hospital admissions which in turn are mainly caused by exacerbation episodes [3].

In an attempt to find a solution to these problems, numerous clinical decision support systems (CDSSs) for the management of patients with COPD have been developed in recent years [4]. In particular, systems based on machine learning algorithms have been developed with the aim of monitoring the health status of patients and foreseeing and preventing exacerbations and hospital admissions [5]. An in-depth research into scientific literature has shown that, in the state of the art, these goals have not yet been met and the

E. Iadanza (✉) · V. A. Mudura
Department of Information Engineering, University of Florence,
Florence, Italy
e-mail: ernesto.iadanza@unifi.it

performance of the current systems is not clinically acceptable. The aim of this work is the design and implementation of a new CDSS that can at least partially fill the current gaps.

2 Materials and Methods

2.1 Data

In order to train, validate and test the decision support system, data from 414 patients affected with COPD and obstructive ventilatory defect were acquired using pulmonary function tests. The following physiological parameters were acquired: Forced Expired Volume in one second (FEV1), Forced Vital Capacity (FVC), Slow Vital Capacity (SVC), FEV1/FVC ratio, FEV1/SVC ratio, Forced Expired Flow at 25–75% (FEF 25–75), Peak Expiratory Flow (PEF), Vital Capacity (VC), Total Lung Capacity (TLC), Residual Volume (RV), Functional Residual Capacity (FRC), Expiratory Reserve Volume, Diffusing Capacity (DLCO), Alveolar Volume (VA) and DLCO/VA ratio. All these

parameters were measured before and after bronchodilation. Other parameters were patients' age, height, bodyweight and sex. According to these parameters five expert pneumologists evaluated the severity of each patient's ventilatory defect and classified it as mild, moderate or severe.

2.2 Data Analysis and Predictive Model Training

Data were processed and analyzed using IBM SPSS Modeler 18.1 [6]. The aim of this phase was to develop a predictive model able to classify the patients' ventilatory defect in three categories (Mild, Moderate and Severe) according to the values of the physiological parameters previously described.

The first step was to try to replicate the performance of support systems for similar decisions, already present in scientific literature. Most of these systems and, in particular, the ones which reached better performances in terms of predictive accuracy, sensitivity and specificity were trained using Neural Networks and Support Vector Machines [7, 8].

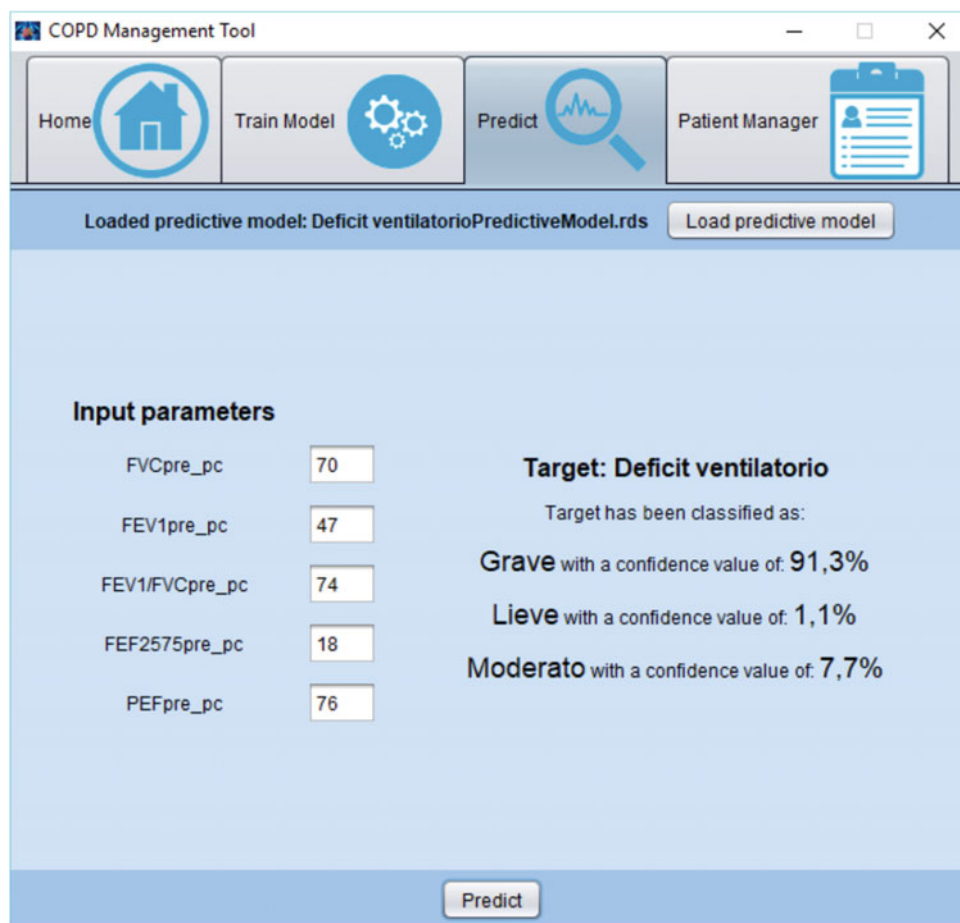


Fig. 1 COPD management tool user interface

Therefore, we used those machine learning techniques to train two different predictive models. We then calculated the performances of these predictive models in predicting the severity of patients' ventilatory defect.

Next step was training a new predictive model with better performances. In order to identify the most suitable machine learning technique for our data, IBM SPSS Modeler's auto classifier node was used. Performances of various machine learning techniques were compared: CART, Random Forest, QUEST, CHAID, Bayesian Networks, Logistic Regression, C5.0, KNN and others. The best performances were reached by the C5.0 algorithm. We therefore concluded that C5.0 was the most suited algorithm for our data. Finally, we trained a predictive model using the C5.0 algorithm and compared its performances with those reached by the predictive models trained with the Neural Network and the Support Vector Machine.

2.3 Predictive Model Implementation

The predictive model trained with the C5.0 algorithm was implemented within a user interface, implemented in Java programming language, the COPD Management Tool. A demonstrative view of the COPD Management Tool is shown in Fig. 1.

3 Results

Results and performances related to the three predictive models, respectively trained using Neural Network, Support Vector Machine and C5.0 algorithms, are reported in the tables below (Tables 1, 2 and 3).

Table 1 Support vector machine performances

Severity	Accuracy	Sensitivity	Specificity
Mild	0.94	0.81	0.97
Moderate	0.89	0.87	0.91
Severe	0.94	0.94	0.94

Table 2 Neural network performances

Severity	Accuracy	Sensitivity	Specificity
Mild	0.93	0.79	0.96
Moderate	0.90	0.86	0.93
Severe	0.97	0.98	0.96

Table 3 C5.0 performances

Severity	Accuracy	Sensitivity	Specificity
Mild	0.94	0.79	0.97
Moderate	0.92	0.91	0.94
Severe	0.97	0.99	0.96

4 Conclusions

Performances obtained with the Neural Network and the Support Vector Machine are comparable with those of the scientific literature. Performances obtained with the C5.0 algorithm are significantly better than those obtained with the two previous model.

The proposed approach, designed with the same systematic approach used in previous works from the authors for a Cardiac Heart Failure CDSS [9–13], allows the evaluation and classification of the results of pulmonary function tests, with excellent performance, compared to the current state of the art and can therefore be used in many clinical applications.

References

1. A. D. Lopez et al., "Chronic obstructive pulmonary disease: current burden and future projections," *Eur. Respir. J.*, vol. 27, no. 2, pp. 397–412, Feb. 2006.
2. T. Vos et al., "Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990? 2015: a systematic analysis for the Global Burden of Disease Study 2015," *Lancet*, vol. 388, no. 10053, pp. 1545–1602, Oct. 2016.
3. S. Ray et al., "The clinical and economic burden of chronic obstructive pulmonary disease in the USA," *Clin. Outcomes Res.*, vol. 5, p. 235, Jun. 2013.
4. D. Sanchez-Morillo et al., "Use of predictive algorithms in-home monitoring of chronic obstructive pulmonary disease and asthma: A systematic review," *Chron. Respir. Dis.*, vol. 13, no. 3, pp. 264–283, 2016.
5. S. C. Peirce et al., "Designing and implementing telemonitoring for early detection of deterioration in chronic disease: Defining the requirements," *Health Informatics J.*, vol. 17, no. 3, pp. 173–190, 2011.
6. "IBM SPSS Modeler." [Online]. Available: <https://www.ibm.com/it-it/marketplace/spss-modeler>. [Accessed: 31-Jan-2018].
7. M. Veezhinathan et al., "Detection of obstructive respiratory abnormality using flow-volume spirometry and radial basis function neural networks.," *J. Med. Syst.*, vol. 31, no. 6, pp. 461–5, Dec. 2007.
8. R. Karakis, I. Guler, and A. H. Isik, "Feature selection in pulmonary function test data with machine learning methods," in *2013 21st Signal Processing and Communications Applications Conference (SIU)*, 2013, pp. 1–4.

9. Guidi, G., Iadanza, E., Pettenati, M.C., Milli, M., Pavone, F., Biffi Gentili, G., "Heart failure artificial intelligence-based computer aided diagnosis telecare system", *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 7251 LNCS, pp. 278–281, 2012.
10. Guidi, G., Pettenati, M.C., Miniati, R., Iadanza, E., "Random forest for automatic assessment of heart failure severity in a telemonitoring scenario" in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, art. no. 6610229, 2013, pp. 3230–3233.
11. Guidi, G., Pettenati, M.C., Melillo, P., Iadanza, E., "A machine learning system to improve heart failure patient assistance", *IEEE Journal of Biomedical and Health Informatics*, 18 (6), art. no. 6851844, pp. 1750–1756, 2014.
12. Guidi, G., Pettenati, M.C., Miniati, R., Iadanza, E., "Heart Failure analysis Dashboard for patient's remote monitoring combining multiple artificial intelligence technologies" in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, art. no. 6346401, 2012, pp. 2210–2213.
13. Guidi, G., Melillo, P., Pettenati, M.C., Milli, M., Iadanza, E., "Performance assessment of a Clinical Decision Support System for analysis of Heart Failure" in *IFMBE Proceedings*, 41, 2014, pp. 1354–1357.