



# An automatic system supporting clinical decision for chronic obstructive pulmonary disease

Ernesto Iadanza<sup>1</sup> · Vlad Mudura<sup>1</sup> · Paolo Melillo<sup>2</sup> · Monica Gherardelli<sup>1</sup>

Received: 18 February 2019 / Accepted: 5 March 2019 / Published online: 16 March 2019  
© The Author(s) 2019

## Abstract

This paper presents a system supporting clinical decisions for patients with Chronic Obstructive Pulmonary Disease (COPD). The system should partially fill the gaps highlighted during an analysis of the current state of the art of Clinical Decision Support Systems (CDSS) for telemonitoring patients affected by COPD. The first step taken was to replicate the performance of similar decision support systems found in the scientific literature. Using physiological parameters drawn from respiratory function tests on 414 patients, two predictive models were created using two machine-learning algorithms: neural network and support vector machine. Performance was comparable to that described in the literature. The results made it possible to affirm that the data available were sufficient to evaluate the extent of respiratory deficit. The next step was to create a new predictive model with better performance than previously obtained. The C5.0 Machine Learning Algorithm was chosen for the development of the model. The resulting performance on the data available was significantly better than with the two previous models. This new predictive model, called COPD, was then implemented in a user interface created using Java programming language. The new software developed, which enables the evaluation and classification of respiratory test results and which can be used in many clinical applications, provides excellent performance compared to the current state of the art.

**Keywords** CDSS · COPD · Home telemonitoring · Machine learning

## 1 Introduction

*Chronic Obstructive Pulmonary Disease (COPD)* is a severe respiratory condition characterized by chronic limitation of airflow. As concerns prevalence and mortality, COPD is classified as the third cause of death in the world [1] and despite mortality appears to be decreasing worldwide [2], COPD still caused more than 3 million deaths in 2010 alone [1]. The Global Burden of Disease Study reports a prevalence of 251 million cases of COPD globally in 2016 [3].

From an economic standpoint, in the United States alone, it is estimated that COPD costs come to about 50 billion dollars

annually. In the coming years, the costs related to the disease are also expected to increase together with its prevalence. Mostly due to the hospitalisation of COPD patients, these costs tend to grow as the disease increases in severity [4]. Hospital admissions are in turn linked mainly to episodes of the exacerbation of COPD's symptoms. Therefore, COPD is characterised by very high prevalence and mortality, considerable healthcare costs mainly due to hospitalization and significant deterioration of the quality of life of the patients affected.

In recent years, to at least partially mitigate and resolve these issues, Clinical Decision Support Systems (CDSS) have been designed and developed for home telemonitoring and management of patients affected by COPD [5]. The main objectives of the CDSS are to monitor and assess the patients' health status and the severity of the disease on a daily basis. This way, the recurrence of the disease can be identified early and, in general, the aggravation of the patient's health state may be prevented [6].

Important outcomes of this monitoring activity are reducing hospital admissions and the resulting costs through early detection and prevention of the exacerbation of the condition with improvement of patient's quality of life. The information

---

✉ Ernesto Iadanza  
ernesto.iadanza@unifi.it

<sup>1</sup> Department of Information Engineering, University of Florence, Via S. Marta 3, 50139 Florence, Italy

<sup>2</sup> Eye Clinic, Multidisciplinary Department of Medical, Surgical and Dental Sciences, University of Campania Luigi Vanvitelli, 80100, Naples, Italy

gathered suggests that, to date, none of these objectives has been fully achieved to any satisfactory extent by any CDSS.

A review of the literature on this issue was undertaken [7–16] and summarized in a recent paper [17]. An analysis of the articles cited, which highlighted the parameters monitored, the algorithms used and the performance (when available), is summarized in Table 1. The physiological parameters in the left column of the table were measured using spirometers, pulse oximeters, blood pressure monitors, scales and thermometers.

Although some studies suggest that telemonitoring has a positive effect on reducing hospital admissions and emergency room (ER) visits, evidence of these benefits is still limited [11, 18–20]. As this analysis indicates, a clinically useful level of accuracy, specificity and sensitivity has not yet been reached to enable the early detection and prevention of the aggravation of the disease. Clearly, appropriate CDSS input parameters and suitable decision-making algorithms still need to be identified.

Another aspect requiring improvement is patient compliance with using the telemonitoring system. Therefore, the system should be minimally invasive, user-friendly and should not restrict the patient's quality of life any more than necessary. These considerations support the need to find a solution to some of these issues and at least partially fill in the gaps found in the current state of the art in this area. Hence, the aim of this research is to design and develop a new CDSS, which has better performance than currently available for the monitoring and management of patients affected by COPD and which can be easily adapted to both hospital and home use.

## 2 Methods and tools

The methodology applied to achieve this goal is summarized in these steps:

- assessment of the adequacy of the data available for classification of the respiratory function test outcomes;

- identification of a new predictive model that outperforms the accuracy, sensitivity and specificity of the models described in the literature;
- implementation of the algorithm software identified in a user interface.

### 2.1 Data

The data used for training and system testing were provided by the Respiratory Physiopathology Department at the Piero Palagi Hospital in Florence. The database comprises 528 anonymized records on 414 patients. There are fewer patients than the number of records because some records were for follow-ups on the same patient, corresponding to medical examinations performed on different dates. Each record contains the measurements of the physiological parameters acquired through the *Respiratory Function Tests* and the *Diffusing Capacity of the Lung for Carbon Monoxide* ( $D_{LCO}$ ) tests. The physiological parameters measured generating data for each patient were: Forced Expired Volume in 1 s ( $FEV_1$ ), Forced Vital Capacity (FVC), Slow Vital Capacity (SVC),  $FEV_1/FVC$  ratio,  $FEV_1/SVC$  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,  $D_{LCO}$ , Alveolar Volume (VA) and  $D_{LCO}/VA$ . All these parameters were measured before and after bronchodilation. Other data taken included each patient's age, height, bodyweight and sex.

Based on the results of the respiratory function and  $D_{LCO}$  tests, five expert pulmonary disease specialists assessed the extent of each patient's respiratory deficits and whether or not their  $D_{LCO}$  was reduced. The respiratory deficits were then classified as mild, moderate or severe. The patients'  $D_{LCO}$  was classified as normal or reduced. During the medical examinations, the patients were also asked if they had experienced any exacerbations, whether they had been admitted to hospital or

**Table 1** State of the art of CDSS from the literature

Telemonitored physiological parameters	Algorithm used for decision making	Performance
- Disease Symptoms	Threshold based methods: these compare the physiological parameters measured with specific thresholds.	Accuracy
- Forced Expired Volume in 1 s ( $FEV_1$ )		40–94%
- Arterial Oxygen	Machine Learning techniques	Specificity
- Saturation ( $SpO_2$ )		70–80%
- Heart rate	<input type="checkbox"/> Neural Network	Sensitivity
- Blood Pressure	<input type="checkbox"/> Bayesian Network	61–95%
- Body weight	<input type="checkbox"/> Pattern Recognition	
- Body temperature	<input type="checkbox"/> CART	
- Respiratory rate	<input type="checkbox"/> Multilevel logistic regression	
	<input type="checkbox"/> Support Vector Machine	

whether they should have had recourse to emergency care due to respiratory problems in the months prior to the examination. Based on this information, the risk of exacerbation was assessed and classified as low or high.

## 2.2 Predictive models

The data discussed above were processed and analysed using IBM SPSS Modeler 18.1 software [21]. This software is a powerful data-mining workbench that enables predictive models to be constructed, without programming. The objective was to create a predictive model based on the input parameters listed above that would enable the classification of the three output parameters of interest: extent of respiratory deficit,  $D_{LCO}$  and risk of exacerbation.

The first step of this phase was to try to replicate the results and performance of similar systems, which used the same input parameters, already in the literature. The two algorithms that achieved the best performance in terms of accuracy, sensitivity and specificity were the Neural Network and the Support Vector Machine (SVM) [22, 23]. Both of these are supervised learning algorithms. They involve a process that, starting from a training dataset, leads to the inference of a mathematical function that links the system's inputs and outputs. The training dataset comprises a series of examples, each of which consists of a known input-output pair. Subsequently, when a new known input is provided, differing from the inputs in the training dataset, the system is able to predict its output, which was not known beforehand. Therefore, the system can generalize the relationships between inputs and outputs learned through the training dataset and then provide outputs from new inputs not considered previously.

The *Neural Network* algorithm builds a network of interconnected nodes organized into layers, which behaves like a network of neurons in the brain. Like a biological neural network, an artificial neural network learns to perform specific tasks through examples it is given. Each node represents an artificial neuron. These are connected by synapses that can transmit a signal from one neuron to another. Each neuron has a state, generally represented by a real number between 0 and 1. Neurons and synapses can also be given a weighting, which varies as training progresses and depending on the examples provided. This weighting will either increase or decrease the strength of the signal being transmitted. The neurons can also have an activation threshold. Based on this parameter, the signal will be transmitted only if its intensity exceeds the threshold level.

*SVM* is a supervised automatic learning technique that treats examples as points in space. Inputs with different outputs are represented by points in space belonging to different regions divided by a clear gap, which is as large as possible. New inputs, which are not part of the training set, are mapped

to one of the different regions of the space, based on how the system was trained.

Therefore, two predictive models were created. The first was based on a neural network and the second on a Support Vector Machine. The two predictive models were used to classify the output as “extent of respiratory deficit”. Performance was calculated for each model in terms of accuracy, sensitivity and specificity. The performance obtained was compared with what was generated by systems already in the literature to verify if the data available were consistent with the proposed objective.

The next step was to develop a new system based on a predictive model that outperformed the two models developed previously. The “Automatic Classifier” function of the IBM SPSS Modeler was used to select the most suitable classification algorithm for the data available. This enabled a rapid comparison of a wide variety of classification algorithms, including CART, Random Forest, QUEST, CHAID, Bayesian Network, Logistic Regression, C5.0, KNN and others. Based on this comparison, the most suitable algorithm to build the predictive model was C5.0, for the data available.

The *C5.0* Algorithm is a supervised automatic learning technique that builds a decision tree for classifying an output parameter starting from a series of examples. Given a set of examples  $S$ , and an output parameter (representing the category to which the various examples belong), the algorithm starts creating the decision tree by performing the following operations:

- If the output of all the examples belonging to  $S$  has the same value and then if all the examples belong to the same class, or if  $S$  is small (there is an internal parameter that defines how small  $S$  can be), the tree will have a single leaf labeled with the most frequent value of the output in  $S$ .
- If the first point is not verified, a test will be defined based on one single input parameter, which will give results in one or more output values. The test represents the root of the tree. There will be a different branch of the tree for each different output value. The  $S$ -set will then be divided into as many different subsets as there are different output values.
- The preceding points are then re-applied recursively to each subset of  $S$ .

To create all the models considered, the operations described below were performed.

### 1) PARTITION OF THE ENTIRE AVAILABLE DATASET INTO A TRAINING SET AND A TEST SET.

This operation is necessary to train and test the predictive model. To identify the optimal percentage of records to

be assigned to each set, 19 different combinations were assessed. For each combination of percentages, the model's accuracy in classifying the output in question on the test set was assessed. Since the partition for a given combination of percentages is random, five different partitions were executed for each combination and the test sets based on the optimal percentages were identified. Average and variance of accuracy of the five partitions were calculated. As an optimal combination of percentages of records to be assigned to training and test sets, the one with the highest average accuracy and the lowest variance of accuracy was selected. The lowest variance of accuracy ensures that the calculated accuracy classification will not depend on, or will only depend minimally on, how the partition is executed. The entire dataset was then partitioned into training and test sets based on the optimal percentages identified.

## 2) TRAINING THE PREDICTIVE MODEL AND CALCULATION OF PERFORMANCE

Once the dataset was optimally divided into training and test sets, the internal parameters of each algorithm were optimized to obtain the best possible performance in terms of accuracy, sensitivity and specificity. Then, the model was trained. Only the training set was used to train the model. Once the training was completed, the model's sensitivity, specificity and accuracy in classifying the output being tested were calculated. Performance was calculated using only the test set. Since test and training set partitions are random, this last step was repeated ten times. The sensitivity, specificity and accuracy of the model were calculated for each partition. The model's final performance was calculated as an average of the performance of the ten partitions. This limited the dependence of the performance on the execution of random partitioning in test and training sets.

## 3 Results

This section shows the performance of the three predictive models applied to the data provided by the Piero Palagi Hospital and illustrates the new CDSS for COPD patient management.

### 3.1 Performance of the different predictive models

Initially, the performance of the predictive models based on the Support Vector Machine and Neural Network algorithms for the extent of respiratory deficit was assessed. Then, based on the C5.0 algorithm, a predictive model of all three outputs of interest (extent of respiratory deficit,  $D_{LCO}$  and risk of exacerbation) was developed and the performance of their sensitivity, specificity and accuracy was calculated. Last, these

performance figures were compared with those obtained from the two previous models.

#### 1) PREDICTIVE MODEL BASED ON THE SUPPORT VECTOR MACHINE ALGORITHM

Table 2 shows the internal parameters of the SVM algorithm used to make the model.

Training used a training set containing 70% of all the records in the entire dataset. The performance achieved by the predictive model in classifying the output as “extent of respiratory deficit” was calculated using a test set containing 30% of the entire dataset. Figure 1 shows the performance for moderate severity degree of this pathology. The performance is compared with those achieved by the models considered below, in similar conditions. Figure 2 shows the performance achieved when mild respiratory deficit is considered. The performance for severe respiratory deficit is shown in Fig. 3.

#### 2) MODEL BASED ON THE NEURAL NETWORK ALGORITHM

Training used a training set containing 85% of all the records in the entire dataset. Performance was calculated using a test set containing 15% of the entire dataset. The type of neural network used for the predictive model was a radial-basis neural network, whose structure is shown in Fig. 4.

Figures 1, 2 and 3 also show the mean performance achieved by the predictive model based on the neural network in classifying the output as “extent of respiratory deficit”, in same way as with the SVM.

#### 3) MODEL BASED ON THE C5.0 ALGORITHM

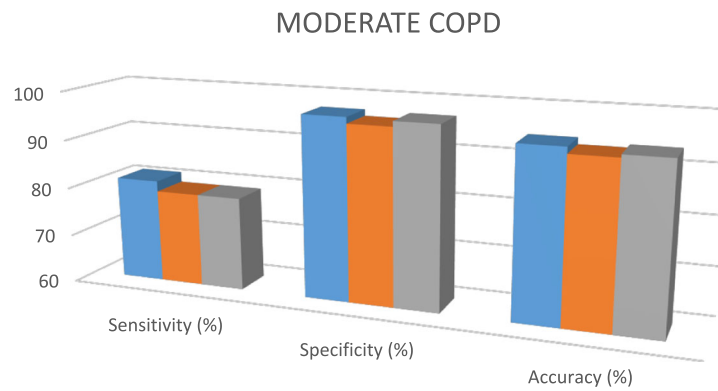
The performance of the predictive model based on the C5.0 algorithm in classifying the extent of respiratory deficit was then evaluated (see Figs. 1, 2 and 3).

In this case, training used a training set containing 80% of all the records in the entire dataset, while performance was calculated using a test set containing 20% of the entire dataset. This predictive model was also used to classify the DLCO. Training used a training set containing 85% of the records of the entire dataset, while performance was calculated using a test set containing 15% of the entire dataset. Performance was 88% for specificity, and 89% for sensitivity and accuracy for normally reduced DLCO levels. Due to the risk of

**Table 2** SVM algorithm internal parameters

Kernel type	RBF
Gamma RBF	0.1
Regularization parameter (C)	10
Regression accuracy ( $\epsilon$ )	0.1

**Fig. 1** Performance comparison of different predictive models in the case of moderate COPD



	Sensitivity (%)	Specificity (%)	Accuracy (%)
■ SVM	81.2	97.2	94.4
■ Neural Network	79.1	95.9	93.1
■ C5.0	79.1	97	93.7

exacerbation, the tests performed showed that the performance of accuracy, sensitivity and specificity do not achieve clinically acceptable levels.

Table 3 shows the internal parameters of the C5.0 algorithm used to make the predictive model.

### 3.2 System implementation

The performance of the predictive model based on C5.0 machine learning algorithm is better than that of other techniques used to interpret and classify the outcome of respiratory function tests. This model was then implemented using an intuitive and easy to use user interface, which enables the operator to make the most of the predictive model’s potential and

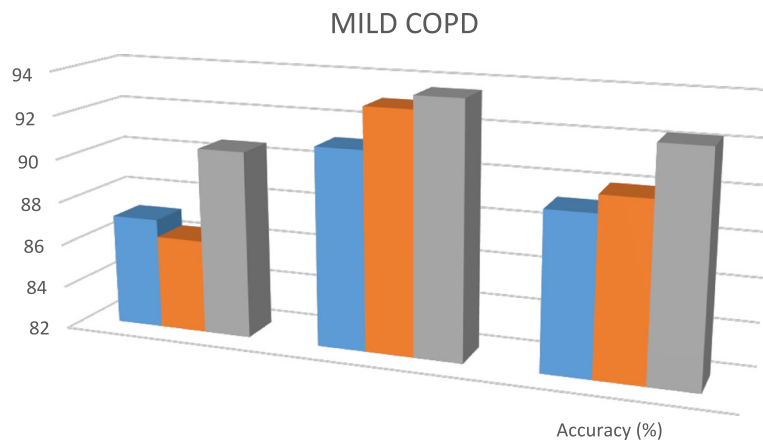
performance without requiring any knowledge of statistics or machine learning by the end user.

The user interface, called *COPD Management Tool*, was created in the Java programming language. The interface has two distinct modules. The first is dedicated to the training process and the creation of the model. The second module is dedicated to the processes of prediction and classification of the respiratory function test results.

#### 1) TRAINING AND CREATION OF THE PREDICTIVE MODEL

The initial screen of the module for training and creation of the predictive model is shown in Fig. 5. The first time the *COPD Management Tool* is used, the user must train and

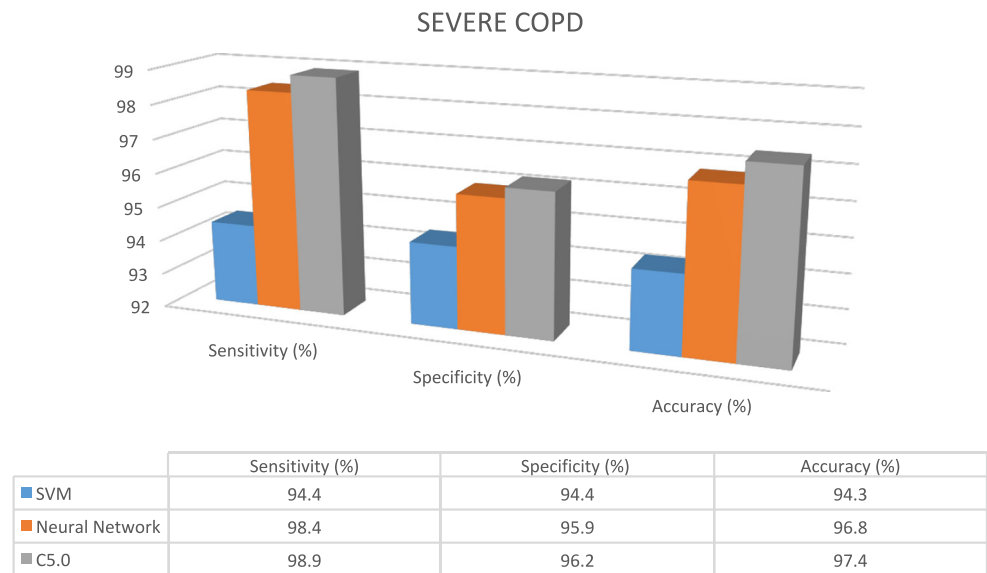
**Fig. 2** Performance comparison of different predictive models in the case of mild COPD



	Sensitivity (%)	Specificity (%)	Accuracy (%)
■ SVM	87.1	91.1	89.2
■ Neural Network	86.3	93	90
■ C5.0	90.6	93.6	92.3



**Fig. 3** Performance comparison of different predictive models in the case of severe COPD



create a predictive model starting from a dataset in an Excel spreadsheet. Therefore, the first step is to load the dataset to be used to train the predictive model. Then, before beginning to train the model, select one of several options available.

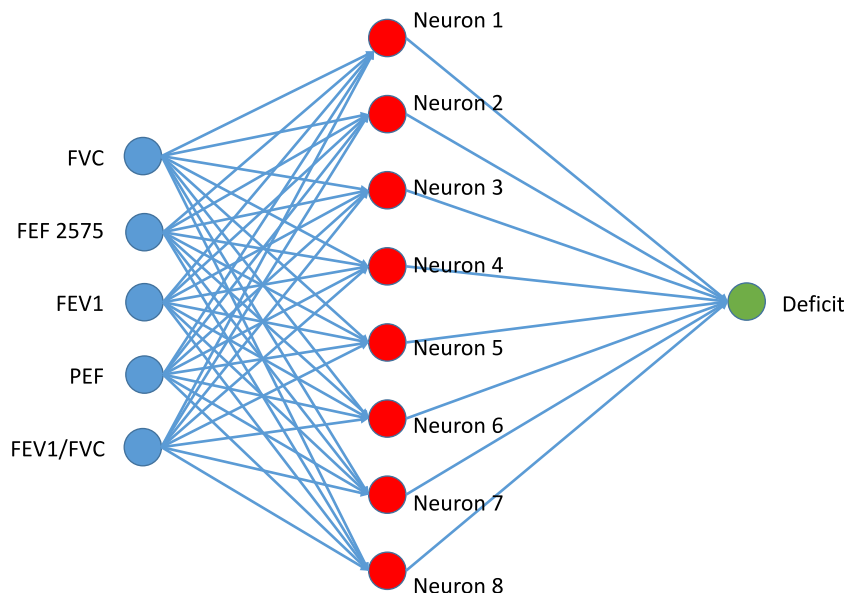
The options are shown in Fig. 6 and described below:

- *Parameter*: the first column lists all the parameters contained in the dataset.
- *Include*: from the second column, the user can choose whether to use the corresponding model training parameter or not.
- *Parameter Type*: in the third column, the user can indicate if the parameter selected will be numeric or categorical.

- *Parameter Role*: from the fourth column, the user can choose whether the parameter in question will belong to the model’s input parameters or if it will be the objective of the prediction and classification process.

Once the desired options have been selected, model training may begin. The model training and creation process includes an optimization phase of the internal parameters of the C5.0 algorithm, which will be performed automatically by the system based on the dataset loaded. The *COPD Management Tool* will then automatically select the optimal parameters for the loaded dataset, meaning that the user need not do this.

**Fig. 4** Structure of a radial-basis neural network



**Table 3** C5.0 algorithm internal parameters

Create split models
Group symbolics
Pruning Severity
Minimum records per child branch
Use global pruning
Winnow attributes
Cost Matrix

## 2) PREDICTION AND CLASSIFICATION OF OUTCOMES

The initial screen of the module for the prediction and classification of outcomes is shown in Fig. 7.

When the *COPD Management Tool* is used for the first time, no predictive model will be loaded. First, the user must load one of the predictive models made with the training module and create a predictive model as described above. Once the desired predictive model is loaded, the following items will be displayed:

- *Input parameters*: these are the parameters that were used previously for training the model and that were selected as input parameters. Now the same parameters can be used to predict and classify the target parameter.
- *Target*: this is the parameter selected previously during the predictive model’s training and creation phase as the objective of the prediction and classification process. This parameter will be classified according to the Input parameter values.

Initially, the system will display a screen in which no prediction or classification of the target parameter has yet been made (Fig. 8). The user can then enter the input parameter values into the appropriate fields. The *COPD Management Tool* will use these values to predict and classify the selected target parameter. Once the input parameter values have been entered, the prediction and classification process can begin. Based on the prediction model trained previously and then loaded and the values of the input parameters entered in their fields, the system will predict and classify the target parameter. In the example in Fig. 9, the target parameter is the extent of respiratory deficit.

The system indicates the different classes to which the target parameter can belong and the confidence level with which the target parameter is assigned to each class. In the example, based on the values of the input parameters entered and the predictive model loaded, the system classifies the extent of respiratory deficit as “Severe”, with a confidence level of 91.3%, “Moderate” with a confidence level of 7.7% and “Mild” with a confidence level of 1%. Actually, the *COPD Management Tool* assesses that the extent of respiratory deficit is severe, based on the values of the input parameters entered.

The fact that the system also provides a measure of the confidence with which it performs a certain classification is a great advantage for the user. This way, the user can assess the reliability of this classification and can decide whether to trust the *COPD Management Tool’s* suggestion or if the analysis should be extended through the performance of another type of evaluation. For example, if the confidence level of a given classification is greater than 90%, it is almost certain that the

**Fig. 5** Structure of a radial-basis neural network

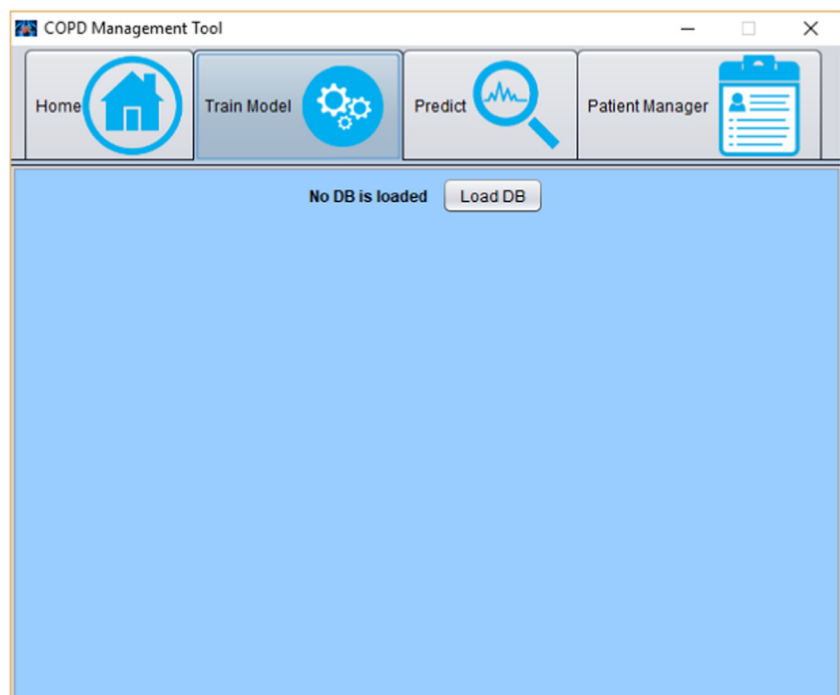
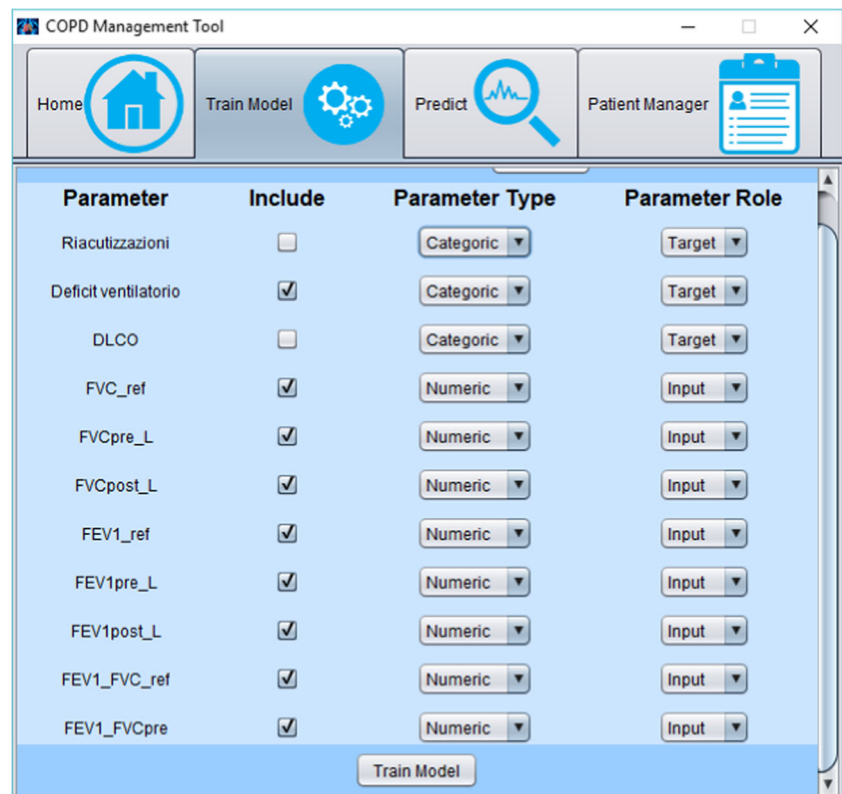


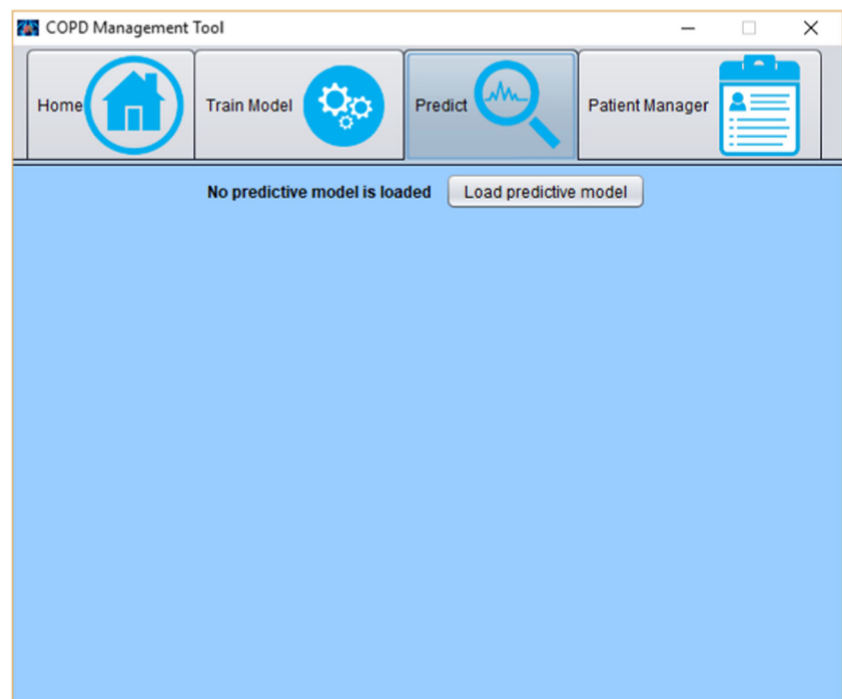
Fig. 6 Options selection screen



classification of the system is correct. If, on the other hand, the classification falls below a certain threshold, the user will know that the classification is not very reliable. Therefore, it would be best to undertake a more detailed analysis and

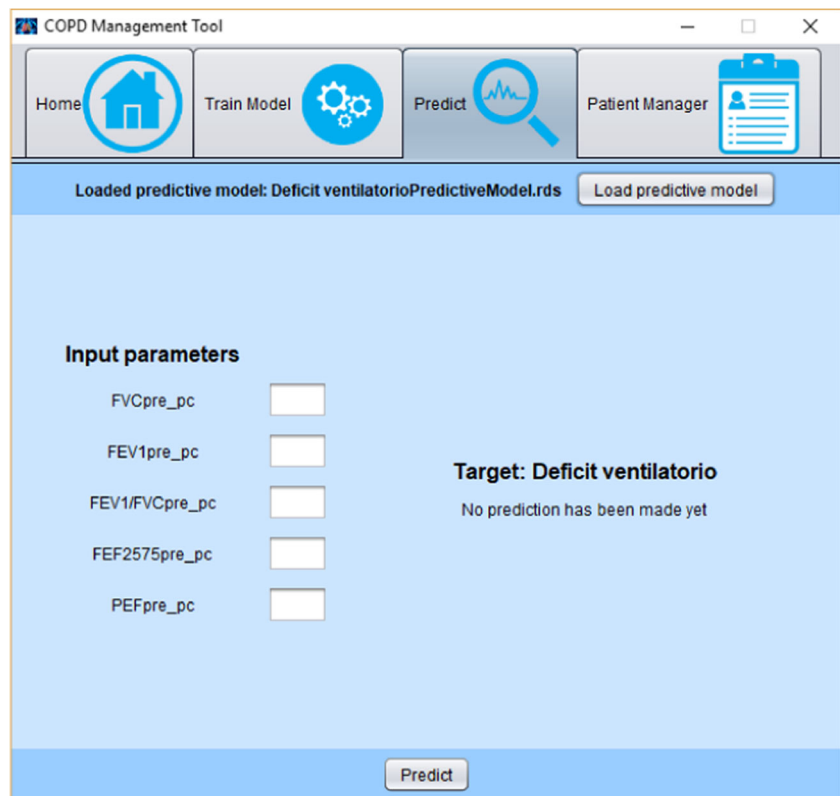
possibly make a different decision. Therefore, the *COPD Management Tool*, as are all decision support systems, is able to efficiently and quickly provide support to the user's decision-making process.

Fig. 7 Initial screen of the module for the prediction and classification of outcomes





**Fig. 8** Target parameter classification and setting screen



**Fig. 9** Target parameter classification



## 4 Discussion of the results

A review of the literature showed that among the systems supporting clinical decisions applied to COPD, neural network and support vector machine algorithms offer the best classification performance in terms of accuracy, specificity and sensitivity. Two predictive models corresponding to these types were created. When classifying the outcomes of respiratory function tests, the performance achieved by the two models using the data available is consistent and comparable with the performance achieved by the systems described in the literature. This means that the data available are suitable for the intended purpose, i.e., the classification of the outcome of the respiratory function tests. However, since the output parameters of the models in the literature do not exactly match the output parameters considered in this paper, a more detailed performance comparison is not possible. In fact, while the objectives of the systems described in the literature were to identify the presence or absence of respiratory obstruction, in this paper the objective is the classification of the extent of the respiratory obstruction and the presence or absence of the reduction of the  $D_{LCO}$ .

After testing the two models cited using the data available, a different algorithm (C5.0) was identified as being more suitable to this dataset for the creation of the predictive model. This predictive model's performance was better than the previous ones. This means that, when interpreting and classifying the results of respiratory function tests, the new predictive model represents a slight yet significant improvement over the current state of the art.

When classifying the risk of exacerbation, the performance of the proposed model is not comparable to the performance of the systems in the literature, because it did not guarantee a clinically acceptable level of accuracy, sensitivity or specificity. The data available were not suitable for predicting and classifying the risk of patients' exacerbation. Therefore, different types of data will be needed to address this problem in future.

To make the most of the potential and performance of the predictive model proposed, a CDSS, called *COPD Management Tool* was developed to enable the end user to make the most of the potential and performance of the predictive model in an easy and intuitive way. Moreover, the user does not need any statistical or machine learning knowledge to use this system.

This instrument can therefore be used by specialist pulmonologists as a tool to support clinical decisions or by general practitioners as a tool to monitor the health of their patients. The most interesting applications are the direct telemonitoring of patients' health status while at home and the generation of alarms in the cases provided for or in the presence of exacerbations. As far as this latter point is concerned, what has already been said in this section should be recalled, i.e. that at this stage, it is not possible to evaluate

performance in a clinically acceptable manner due to the inadequacy of the data.

## 5 Conclusion

A detailed review of COPD and the clinical-decision support systems specifically applied to COPD is described in the literature. Specifically, a study was made of the algorithms most widely used to create predictive models on which current decision support systems are based. Then the physiological parameters most used to train these models to make predictions on the target parameters were identified.

Once these elements were identified, an attempt was made to replicate the performance, in terms of sensitivity, specificity and accuracy, of the two systems found in the literature: the Neural Network and the Support Vector Machine. Subsequently, a new predictive model based on C5.0 machine learning algorithm was developed. The C5.0 algorithm gave better performance than the two previous models.

The proposed system, designed using the same approach applied in previous research by the authors on Heart Failure CDSS [24–28], enables the evaluation and classification of the results of pulmonary function tests, with good performance, compared to the current state of the art. Therefore, this system can be used in many clinical applications.

The *COPD Management Tool*, which is a CDSS based on the C5.0 machine-learning algorithm, was developed to enable the interpretation and classification of the results of respiratory functional tests [29]. This instrument offers significant improvements in performance compared to the current state of the art in this sector. Doctors can apply this instrument to many contexts, including home telemonitoring. In fact, home use allows continuous monitoring of patients' health status and the generation of alarms in the cases provided for. In future, the *COPD Management Tool* can also be applied to monitoring and management of diseases other than COPD. Actually, the processes that led to the tool's design and implementation, including the C5.0 algorithm, are generalised and unrelated to any particular condition. Potentially, they can be applied to any type of data and to the management of any type of disease.

**Acknowledgments** The authors wish to express their thanks to Dr. Walter Castellani, Dr. Marco Chiostrì and Dr. Laura Frilli, and all the Medical Staff of the Azienda USL Toscana Centro – Ospedale Piero Palagi in Firenze, for their contribution to this study.

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

**Ethical approval** This article does not contain any studies with human participants or animals performed by any of the authors.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Lozano R, Naghavi M, Foreman K, Lim S, Shibuya K, Aboyans V, et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the global burden of disease study 2010. *Lancet*. 2012;380:2095–128.
- Lopez-Campos JL, Ruiz-Ramos M, Soriano JB. Mortality trends in chronic obstructive pulmonary disease in Europe, 1994–2010: a joinpoint regression analysis. *Lancet Respir Med*. 2014;2:54–62.
- Vos T, Allen C, Arora M, Barber RM, Bhutta ZA, Brown A, 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*. 2016;388(10053):8–14. [https://doi.org/10.1016/S0140-6736\(16\)31678-6](https://doi.org/10.1016/S0140-6736(16)31678-6).
- Guarascio SM, Ray SM, Finch CK, Self TH. The clinical and economic burden of chronic obstructive pulmonary disease in the USA. *Clinicoecon and Outcomes Res*. 2013;2015(5):235–45. <https://doi.org/10.2147/CEOR.S34321>.
- Sanchez-Morillo D, Fernandez-Granero MA, Leon-Jimenez A. Use of predictive algorithms in-home monitoring of chronic obstructive pulmonary disease and asthma: a systematic review. *Chron Respir Dis*. 2016;13(3):264–83. <https://doi.org/10.1177/1479972316642365>.
- Peirce SC, Hardisty AR, Preece AD, Elwyn G. Designing and implementing telemonitoring for early detection of deterioration in chronic disease: defining the requirements. *Health Informatics J*. 2011;17(3):173–90. <https://doi.org/10.1177/1460458211409717>.
- Koff PB, Jones RH, Cashman JM, Voelkel NF, Vandivier RW. Proactive integrated care improves quality of life in patients with COPD. *Eur Respir J*. 2009;33(5):1031–8. <https://doi.org/10.1183/09031936.00063108>.
- Sund ZM, Powell T, Greenwood R, Jarad NA. Remote daily real-time monitoring in patients with COPD - a feasibility study using a novel device. *Respir Med*. 2009;103(9):1320–8. <https://doi.org/10.1016/j.rmed.2009.03>.
- Jensen MH, Cichosz SL, Dinesen B, Hejlesen OK. Moving prediction of exacerbation in chronic obstructive pulmonary disease for patients in telecare. *J Telemed Telecare*. 2012;18(2):99–103. <https://doi.org/10.1258/jtt.2011.110607>.
- Yañez AM, Guerrero D, Pérez De Alejo R, Garcia-Rio F, Alvarez-Sala JL, Calle-Rubio M, et al. Monitoring breathing rate at home allows early identification of COPD exacerbations. *Chest*. 2012;142(6):1524–9. <https://doi.org/10.1378/chest.11-2728>.
- Pinnock H, Hanley J, McCloughan L, Todd A, Krishan A, Lewis S, et al. Effectiveness of telemonitoring integrated into existing clinical services on hospital admission for exacerbation of chronic obstructive pulmonary disease: researcher blind, multicenter, randomized controlled trial. *BMJ*. 2013;347:f6070. <https://doi.org/10.1136/bmj.f6070>.
- Fernández-Granero MA, Sánchez-Morillo D, León-Jiménez A, Crespo LF. Automatic prediction of chronic obstructive pulmonary disease exacerbations through home telemonitoring of symptoms. *Bio-Med Mater Eng*. 2014;24(6):3825–32. <https://doi.org/10.3233/BME-141212>.
- Burton C, Pinnock H, McKinstry B. Changes in telemonitored physiological variables and symptoms prior to exacerbations of chronic obstructive pulmonary disease. *J Telemed Telecare*. 2015;21(1):29–36. <https://doi.org/10.1177/1357633X14562733>.
- Hardinge M, Rutter H, Velardo C, Shah SA, Williams V, Tarassenko L, et al. Using a mobile health application to support self-management in chronic obstructive pulmonary disease: a six-month cohort study. *BMC Medical Informatics and Decision Making*. 2015;15(46):46. <https://doi.org/10.1186/s12911-015-0171-5>.
- Mohhtar MS, Redmond SJ, Antoniadis NC, Rochford PD, Pretto JJ, Basilakis J, et al. Predicting the risk of exacerbation in patients with chronic obstructive pulmonary disease using home telehealth measurement data. *Artif Intell Med*. 2015;63(1):51–9. <https://doi.org/10.1016/j.artmed.2014.12.003>.
- Sanchez D, Miguel M, Fernandez A. Detecting COPD exacerbations early using daily telemonitoring of symptoms and k-means clustering : a pilot study. *Med Biol Eng Comput*. 2015;53(5):441–51. <https://doi.org/10.1007/s11517-015-1252-4>. Accessed on 28 Aug 2018.
- Mudura VA, Frosini F, Iadanza E. Clinical decision support systems for COPD: a general overview. In: Eskola H, Väisänen O, Viik J, Hyttinen J, editors. *EMBE & NBC 2017. EMBEC 2017, NBC 2017. IFMBE proceedings*, vol. 65. Singapore: Springer; 2018. [https://doi.org/10.1007/978-981-10-5122-7\\_234](https://doi.org/10.1007/978-981-10-5122-7_234).
- Pedone C, Lelli D. Systematic review of telemonitoring in COPD: an update. *Pneumonol Alergol Pol*. 2015;83(6):476–84. <https://doi.org/10.5603/PiAP.2015.0077>.
- Cruz J, Brooks D, Marques A. Home telemonitoring effectiveness in COPD: a systematic review. *Int J Clin Pract*. 2014;68(3):369–78. <https://doi.org/10.1111/ijcp.12345>.
- Gurbeta L, Badnjevic A, Maksimovic M, Omanovic-Miklicanin E, Sejdic E. A telehealth system for automated diagnosis of asthma and chronic obstructive pulmonary disease. *J Am Med Inform Assoc*. 2018;25(9):1213–7. <https://doi.org/10.1093/jamia/ocy055.2018>.
- IBM: IBM SPSS Modeler [online]. Available: <https://www.ibm.com/it-it/marketplace/spss-modeler>. Accessed on 28 Aug 2018.
- Veezhinathan M, Ramakrishnan S. Detection of obstructive respiratory abnormality using flow-volume spirometry and radial basis function neural networks. *J Med Syst*. 2007;31(461):461–5. <https://doi.org/10.1007/s10916-007-9085-9>.
- Karakis R, Guler I, Isik AH. Feature selection in pulmonary function test data with machine learning methods. In: *Proc. 2013 21st signal processing and communications applications conference (SIU), Haspolat*; 2013. p. 1–4. <https://doi.org/10.1109/SIU.2013.6531578>.
- Guidi G, Iadanza E, Pettenati MC, Milli M, Pavone F, Biffi Gentili G. Heart failure artificial intelligence-based computer aided diagnosis telecare system. In: Donnelly M, Paggetti C, Nugent C, Mokhtari M, editors. *Impact analysis of solutions for chronic disease prevention and management. ICOST 2012. Lecture notes in computer science*, vol. 7251. Berlin: Springer; 2012. [https://doi.org/10.1007/978-3-642-30779-9\\_44](https://doi.org/10.1007/978-3-642-30779-9_44).
- Guidi G, Pettenati MC, Miniati R, Iadanza E. Random forest for automatic assessment of heart failure severity in a telemonitoring scenario. In: *Proc. 2013 35th annual international conference of the IEEE engineering in medicine and biology society (EMBC)*, Osaka; 2013. p. 3230–3. <https://doi.org/10.1109/EMBC.2013.6610229>.
- Guidi G, Pettenati MC, Melillo P, Iadanza E. A machine learning system to improve heart failure patient assistance. *IEEE J Biomed Health Inform*. 2014;18(6):1750–6. <https://doi.org/10.1109/JBHI.2014.2337752>.
- Guidi G, Pettenati MC, Miniati R, Iadanza E. Heart failure analysis dashboard for patient's remote monitoring combining multiple artificial intelligence technologies. In: *Proc. 2012 annual international*

- conference of the IEEE engineering in medicine and biology society, San Diego, CA; 2012. p. 2210–3. <https://doi.org/10.1109/EMBC.2012.6346401>.
28. Guidi G, Melillo P, Pettenati MC, Milli M, Iadanza E. Performance assessment of a clinical decision support system for analysis of heart failure. In: Roa RL, editor. XIII Mediterranean conference on medical and biological engineering and computing 2013. IFMBE proceedings, vol. 41. Cham: Springer; 2014. [https://doi.org/10.1007/978-3-319-00846-2\\_335](https://doi.org/10.1007/978-3-319-00846-2_335).
29. Iadanza E, Mudura VA. A decision support system for chronic obstructive pulmonary disease (COPD). In: Lhotska L, Sukupova L, Lacković I, Ibbott G, editors. World congress on medical physics and biomedical engineering 2018. IFMBE proceedings, vol 68/3. Singapore: Springer; 2019. [https://doi.org/10.1007/978-981-10-9023-3\\_57](https://doi.org/10.1007/978-981-10-9023-3_57).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.