



## An artificial intelligence approach to early predict symptom-based exacerbations of COPD

Miguel Angel Fernandez-Granero, Daniel Sanchez-Morillo & Antonio Leon-Jimenez

**To cite this article:** Miguel Angel Fernandez-Granero, Daniel Sanchez-Morillo & Antonio Leon-Jimenez (2018) An artificial intelligence approach to early predict symptom-based exacerbations of COPD, *Biotechnology & Biotechnological Equipment*, 32:3, 778-784, DOI: [10.1080/13102818.2018.1437568](https://doi.org/10.1080/13102818.2018.1437568)

**To link to this article:** <https://doi.org/10.1080/13102818.2018.1437568>



© 2018 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 10 Feb 2018.



Submit your article to this journal [↗](#)



Article views: 5186



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 22 View citing articles [↗](#)

## An artificial intelligence approach to early predict symptom-based exacerbations of COPD

Miguel Angel Fernandez-Granero<sup>a</sup>, Daniel Sanchez-Morillo<sup>a</sup> and Antonio Leon-Jimenez<sup>b</sup>

<sup>a</sup>Biomedical Engineering and Telemedicine Lab, School of Engineering, University of Cádiz, Cádiz, Spain; <sup>b</sup>Pulmonology and Allergy Unit, Puerta del Mar University Hospital, Cádiz, Spain

### ABSTRACT

Acute exacerbations are one of the main causes that reduce health-related quality of life and lead to hospitalisations of patients of chronic obstructive pulmonary disease (COPD). Prediction of exacerbations could diminish those negative effects and reduce the high costs associated with COPD patients. In this study, 16 patients were telemonitored at home during six months. Respiratory sounds were recorded daily with an electronic sensor ad-hoc designed. In order to enable an automatic prediction of symptom-based exacerbations, recorded data were used to train and validate a decision tree forest classifier. The developed model was capable of predicting early acute exacerbations of COPD, as average, with a 4.4 days margin prior to onset. Thirty-two out of 41 exacerbations were detected early. A percentage of 75.8% (25 out of 33) of detected episodes were reported exacerbation and 87.5% (7 out of 8) were unreported events. The achieved results demonstrated that machine-learning techniques have significant potential to support the early detection of COPD exacerbations.

### ARTICLE HISTORY

Received 16 February 2017  
Accepted 3 February 2018

### KEYWORDS

COPD; acute exacerbation; telehealth; respiratory sounds; early detection; prediction; telemonitoring

### Introduction

Chronic obstructive pulmonary disease (COPD) is ranked as the third most common global cause of death in the last decade, being a main cause of chronic morbidity [1]. Acute exacerbations of COPD (AECOPD) are one of the prime causes that reduce health-related quality of life and lead COPD patients to hospitalisations. Early detection of exacerbations, achieved with the support of telehealth systems and interventions aimed to promote self-management, could decrease the negative effects and costs associated with COPD patients.

A COPD self-management intervention has been defined by an International Expert Group consensus [2] as a 'structured but personalised and often multi-component, with goals of motivating, engaging and supporting the patients to positively adapt their health behaviour and develop skills to better manage their diseases.'

Home telehealth systems are being designed and developed to help patients with chronic disorders, carers and health and social professionals. Telehealth approaches can contribute to avoiding the need for frequent visits to hospital and provide an easy, secure and specific tool to self-manage and control the disease [3–7]. These telehealth systems are usually composed of sensors and devices for obtaining health related biomarkers.

A recent Cochrane review has described potential advantages of self-management interventions in COPD [8]. In addition, a more recent systematic review has reported that predictive data mining models with good clinical reliability are an important goal for the future development of telehealth and self-management in chronic respiratory conditions [9].

COPD symptoms often change due to exacerbations [10]. However, the AECOPD definition is controversial, mainly because of its heterogeneous nature [11,12]. Two definitions of COPD exacerbations are widely used in most studies: a) an event-based episode is defined as an attendance in an emergency or primary care setting with worsening of symptoms, and, in some studies, including the self-medication of patients with a self-management plan with corticosteroids and/or antibiotics; and b) a symptom-based exacerbation characterized by the increasing of respiratory symptoms [12]. The decision on the selected definition shapes the number of exacerbations observed and the algorithm efficiency [12]. In summary, the performance of algorithms may be adversely affected for the lack of agreement on a definition of an exacerbation [13] and by unreported AECOPD [14]. Consequently, intelligent algorithms must be robust notwithstanding the definition of exacerbation of COPD applied.

This study aims at exploring the performance of daily home recorded respiratory sounds for early detection of symptom-based AECOPD by using computerized analysis and artificial intelligence techniques.

## Subjects and methods

### Subjects

Sixteen COPD patients participated in this study. The participants were retrieved in the Pulmonology and Allergy Department of Puerta del Mar University Hospital, Cadiz, Spain. The patients had been diagnosed with COPD by spirometry and had been classified into GOLD groups D and C [15]. The patients were  $70.2 \pm 6.6$  years old and had a history of two or more exacerbations requiring treatment with antibiotics or oral corticosteroids or of one hospital admission with an exacerbation in the previous year and cumulative tobacco consumption of  $>20$  pack-years. Informed consent was signed by all participants and the study obtained an ethics approval from the local Ethics Committee. The patients used a sensor device and a base station to record their respiratory sounds daily for 6 months [16]. Respiratory sounds were recorded on the suprasternal notch following the instructions presented by a multimodal interface running in the base-station. Daily recordings were sent to the central server located at the Hospital and added to the electronic patient record.

### Features extraction: discrete wavelet transform

In general, it is accepted that the frequency of respiratory sounds is in the range of 100–2500 Hz, and that

tracheal sounds can reach 4000 Hz. The spectrum of heart sounds is located between 20 and 100 Hz [17]. Respiratory sounds are non-stationary, and consequently, they are suitable for time frequency analysis using discrete wavelet transform (DWT) [18].

Selection of the maximum wavelet decomposition level depends on dominant frequencies of the signal. In this study, two levels for wavelet decomposition were chosen, since significant frequency range of respiratory sounds signals is 100–2000 Hz.

The wavelet features were obtained on the sub-bands A1, A2 and D2 (Figure 1). Biorthogonal 1.5 wavelet function was used [19,20]. From each of the three sub-bands, the following statistical parameters were estimated:

- (1) Mean of the absolute values of the wavelet coefficients;
- (2) Average power and standard deviation of the wavelet coefficients;
- (3) Ratio of the absolute mean values of coefficients adjacent sub-bands;
- (4) Kurtosis and skewness values of wavelet coefficients.

In total, 18 DWT features were extracted. Further details can be found in [16].

### Fast correlation-based filter

Feature subset selection (FSS) aims to remove certain predictor variables that may become unimportant or even be redundant. Fast correlated-based filter (FCBF) [21] is a multivariable algorithm to measure the class-feature and

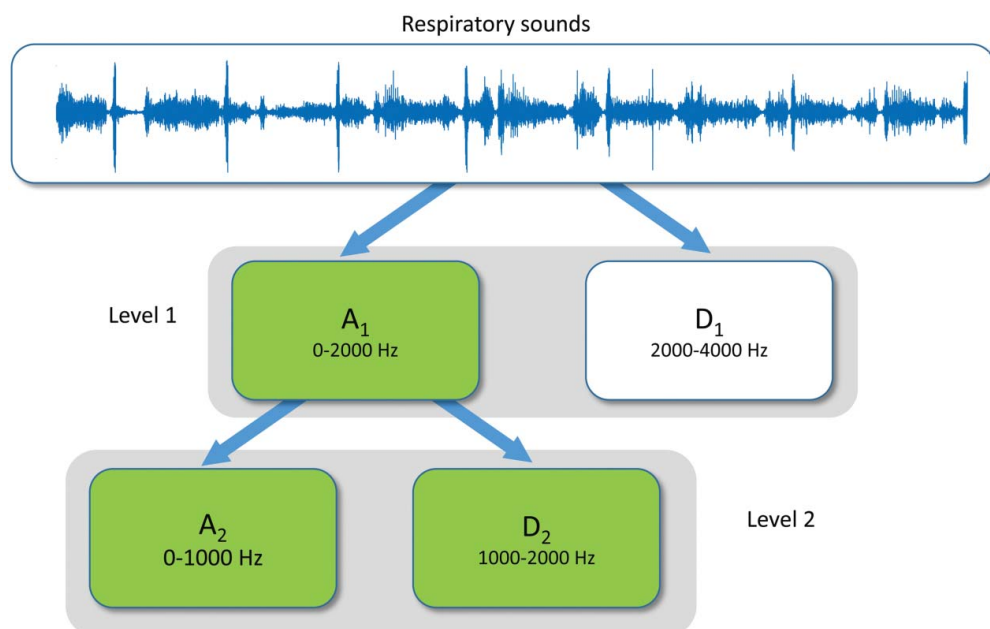


Figure 1. Sub-bands used in the discrete wavelet transform implementation (marked in green colour).

the feature–feature correlation. FCBF is used with high-dimensional data. It has been demonstrated to be effective in meeting the goals of FSS [22]. FCBF begins choosing a set of features strongly correlated with the class based on the symmetric uncertainty (SU). SU is given by:

$$SU(X; Y) = \frac{IG(X; Y)}{H(X) + H(Y)} \quad (1)$$

where  $IG(X; Y)$  is the gain information and  $H(X)$ ,  $H(Y)$  are the individual entropies. FSS for classification is a method that recognises every predominant feature to the class concept and eliminates all others. The method fully explained in [21] was followed where the three suggested heuristics were applied in this study. They were selected because they can successfully recognise predominant features and eliminate redundant ones among all significant features, avoiding pairwise exploration of F-correlations.

### Decision tree forest

A decision tree forest (DTF) is an assembly of decision trees with combined predictions to perform the global prediction for the forest. It is well known that the ensemble methods can be used to improve the performance prediction [23]. A DTF produces a number of parallel independent trees. The ‘out of bag’ data rows were used for validation of the decision tree forest model. This method provides an independent test without demanding a separate dataset [24].

### Results and discussion

The symptom-based definition of AECOPD was used in this study. An AECOPD was defined according to the

Anthonisen criteria [25,26]. Markov Chain Monte Carlo (MCMC) function was applied to impute missing data were IBM® SPSS was used [27,28].

The input dataset included the 18 wavelet features on the sub-bands,  $A_1$ ,  $A_2$  and  $D_2$ . The FSS method was applied for features reduction using FCBF. As a result, 11 features were retained. Average power in each band, mean of absolute values in bands A1 and A2 and standard deviations in bands A1 and D2 were excluded.

A decision tree forest classifier was then designed for the early prediction of AECOPD [29,30]. The number of trees was selected using a minimum error cost function. The resulting forest had 40 trees (error cost = 0.0708). Three predictors (out of 11) were selected for each split. The maximum depth of a tree in the forest was 26.

Receiver operating characteristic (ROC) analysis was carried out [31] (Figure 2). The performance was evaluated according to accuracy, specificity, sensitivity, confusion matrix, positive and negative predictive values, F1 recall and Matthews correlation coefficient. MathWorks MATLAB® software was used.

Respiratory symptoms during the prodromal phase of a COPD exacerbation may get worse for 7 days prior to the onset [32]. Accordingly, a categorical dichotomous variable was used to define the target. The classifier output targeted a decision rule for reducing the number of false alarms. This decision rule defined that an AECOPD was established if, for two or more successive days, patients suffered an increase of respiratory symptoms [33]. Consequently, an alarm was set if a positive output was produced in the classifier for two successive days. Supplementary information about this method can be found in [16] and [34].

Finally, 15 patients completed the study and one patient was excluded (Figure 3). During the six months

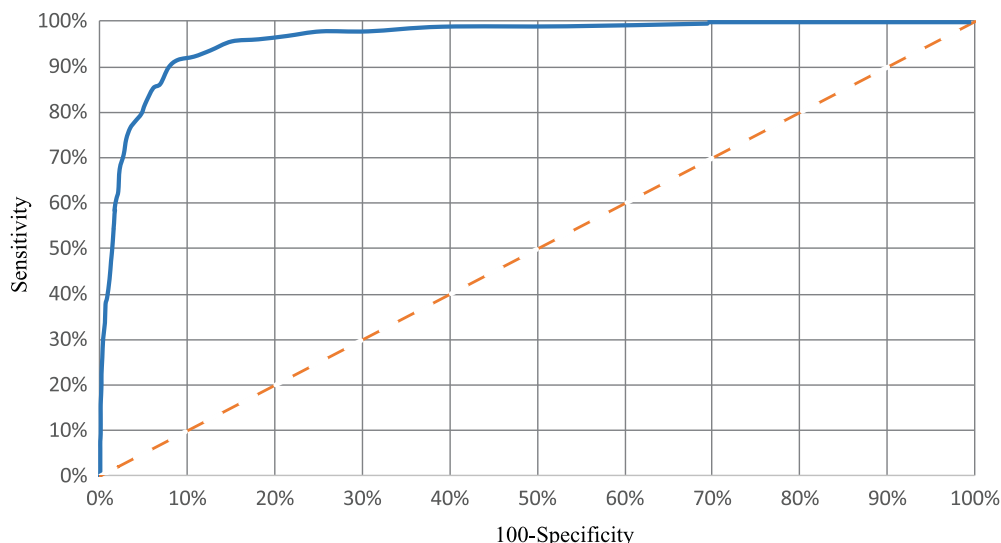
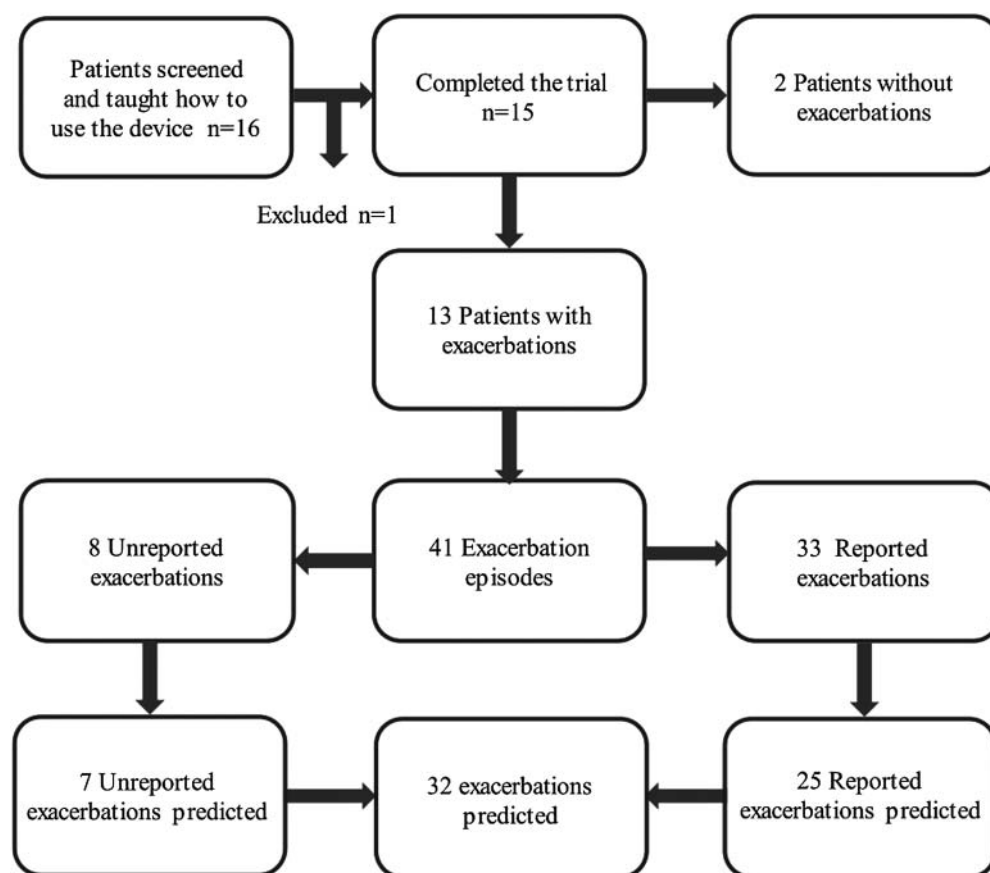


Figure 2. Receiver operating characteristic curve for the validated DTF classifier.



**Figure 3.** Flowchart with complete information on patient involvement, dropout and AECOPD predicted during the pilot study using a decision tree forest. Symptom-based exacerbations were considered.

of home monitoring, the patients suffered 41 episodes that matched the Anthonisen criteria and therefore were considered as symptom-based AECOPD. Thirty-three out of the 41 episodes were reported episodes. In reported episodes, specific medical attention was required and annotated in the patients' health records. The remaining 8 out of 41 AECOPD were not reported episodes that were detected through symptoms monitoring. In these episodes, the patients did not seek medical assistance.

A total of 2104 days, each represented by 11 parameters were used to train and validate the DTF classifier.

Table 1 shows some performance parameters of the validated classifier: 78% (32 out of 41) of AECOPD were detected early with a margin of  $4.4 \pm 1.8$  days prior to the day in which the patients met the Anthonisen criteria. In addition, 75.8% (25 out of 33) were reported AECOPD and 87.5% (7 out of 8) were unreported AECOPD with a prediction margin of  $4.5 \pm 1.5$  and  $4 \pm 2$  days respectively. Two false positives were registered.

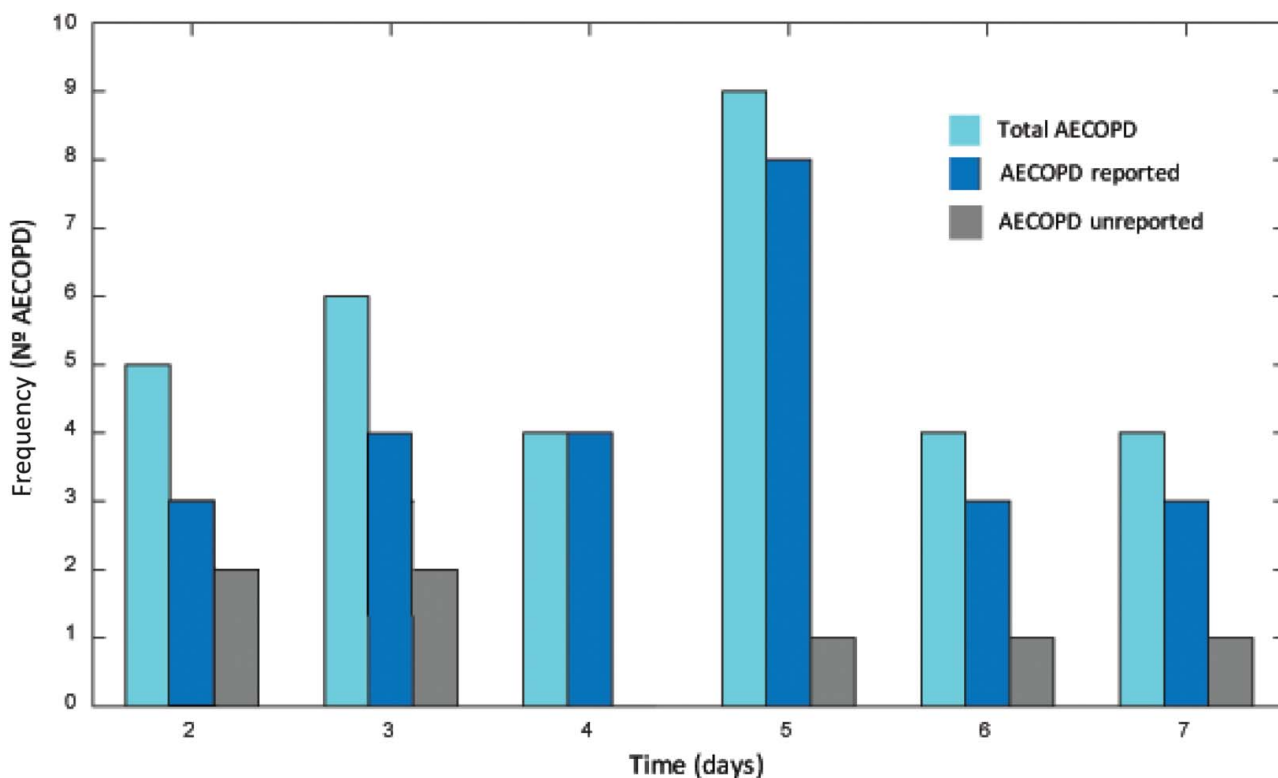
Figure 4 shows the histogram of the prediction margins of the total, reported and unreported AECOPD, for the validated DTF classifier and the applied decision rule.

Finding predictors with clinical consistency remains a priority for the development of interventions of domiciliary telemonitoring in COPD [35]. However, only a few studies have been published on early detection of acute exacerbation of COPD based only on physiological measurements.

In a recent study, computerised analysis of respiratory sounds has been demonstrated potentiality as a consistent indicator of respiratory status in patients with COPD

**Table 1.** Classifier performance evaluation.

Indicator	Value
Total AECOPD predicted	32 (78.0%)
AECOPD reported predicted	25 (75.8%)
AECOPD unreported predicted	7 (87.5%)
True positives	35.6%
False positives	2.2%
True negatives	52.2%
False negatives	10.0%
Accuracy	87.8%
Sensitivity	78.1%
Specificity	95.9%
Positive predictive value	94.1%
Negative predictive value	83.9%
False alarms of the system	2
F1 score	0.8
Matthews correlation coefficient	0.8



**Figure 4.** Histogram of prediction margins of total, reported and unreported AECOPD. The horizontal axis indicates the days of prediction prior to AECOPD onset.

[36]. In the present study, the feasibility of respiratory sounds to early predict symptom-based AECOPD was explored. Although, in this study, the exacerbations were established according to the Anthonisen criteria (symptom-based exacerbation), the results achieved in this study and in prior works, with the same cohort of patients, are promising and consistent with the two most relevant definitions of exacerbations.

Using respiratory sounds and event-based AECOPD, the authors evaluated a model with a similar predictive capacity. This way, 75.8% AECOPD were predicted with an average of 5 days in advance of medical intervention [16]. Predictive models that used symptoms acquired by a multimodal home base station to detect event-based AECOPD [34] and symptoms-based episodes [26] have also been reported by authors.

In other published studies, symptoms and physiological parameters have been explored for the detection of AECOPD. In [37], prediction of risk of exacerbation in the following 30 days was assessed using linear discriminant functions, with a sensitivity of 70%. Bayesian network models were used in [38] and a true positive rate of 0.88 was obtained. Multilevel logistic regression was explored in [39] but physiological variables seemed not to differentiate between exacerbations and isolated bad days. Finally, CART algorithms were used in [40], to classify

home telehealth measurement data into risks categories with a sensitivity of 61.1%.

Prediction algorithms must be robust regardless of the exacerbation of COPD definition applied. The definition of AECOPD affects the number of detected episodes and therefore could result in biased outcomes in the performance of the algorithm used for early detection [12]. For this reason, the suggested system has demonstrated good prediction outcomes and robustness against the two most used AECOPD definitions using both symptoms and respiratory sounds as predictors.

## Conclusions

A DTF classifier was designed using features extracted from the respiratory sounds daily recorded at home with a respiratory sensor. The proposed system was able to predict symptom-based episodes with 4.4 days of margin, as average, prior to onset. The detection accuracy was 78.0% and 75.8% of the detected episodes were reported exacerbations, whereas 87.5% were unreported events. The proposed method may automatically enable patients to request medical attention or to initiate a self-management plan. A larger sample of patients and a longer period could support in defining whether the results described in this work are generalizable. The results

obtained in the present work suggest that the described methodology and the designed electronic sensor could aid the design of consistent intelligent algorithms aimed at predicting AECOPD and consequently could provide support both to patients and physicians.

### Disclosure statement

The authors declare no conflict of interest.

### Funding

This work was supported in part by the Ambient Assisted Living (AAL) E.U. Joint Programme, by grants from Ministerio de Educación y Ciencia of Spain and Instituto de Salud Carlos III [grant number PI08/90946] and [grant number PI08/90947].

### References

- [1] Patel J, Burney PG, Newson RB, et al. Global and regional trends in mortality from chronic obstructive pulmonary disease: their relation to poverty, smoking and population change. *Eur Respir J*. 2014;44(Suppl 58):421.
- [2] Effing TW, Vercoulen JH, Bourbeau J, et al. Definition of a COPD self-management intervention: International Expert Group consensus. *Eur Respir J*. 2016;48:46–54.
- [3] Giorginoll T, et al.. The HOMEY project: a telemedicine service for hypertensive patients personalisation for e-Health. In: Floriana G, et al., editors. Proceedings from the 1st International Workshop on Personalisation for e-Health held in conjunction with UM05; 2005 Jul 29; Edinburgh, Scotland. Sheffield (UK): The University of Sheffield; 2005. p. 21–30.
- [4] Young M, Sparrow D, Gottlieb D, et al. A telephone-linked computer system for COPD care. *Chest*. 2001;119:1565–1575.
- [5] Mooney KH, Beck SL, Friedman RH, et al. Telephone-linked care for cancer symptom monitoring: a pilot study. *Cancer Pract*. 2002;10:147–154.
- [6] Pinto BM, Friedman R, Marcus BH, et al. Effects of a computer-based, telephone-counseling system on physical activity. *Am J Prev Med*. 2002;23:113–120.
- [7] Ramelson HZ, Friedman RH, Ockene JK. An automated telephone-based smoking cessation education and counseling system. *Patient Educ Couns*. 1999;36:131–144.
- [8] Zwerink M, Brusse-Keizer M, van der Valk PD, et al. Self management for patients with chronic obstructive pulmonary disease. *The Cochrane Library*. 2014 [cited 2018 Jan 31];3:CD002990. DOI: 10.1002/14651858.CD002990.pub3.
- [9] 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):264–283.
- [10] Vestbo J, Hurd SS, Agustí AG, et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. *Am J Respir Crit Care Med*. 2013;187(4):347–365.
- [11] Mackay AJ, Donaldson GC, Patel AR, et al. Detection and severity grading of COPD exacerbations using the exacerbations of chronic pulmonary disease tool (EXACT). *Eur Respir J*. 2014;43(3):735–744.
- [12] Effing TW, Kerstjens HA, Monninkhof EM, et al. Definitions of exacerbations: does it really matter in clinical trials on COPD? *Chest*. 2009;136:918–923.
- [13] McKinstry B, Pinnock H, Sheikh A. Telemedicine for management of patients with COPD? *The Lancet*. 2009;374(9691):672–673.
- [14] Trappenburg JC, Touwen I, de Weert-van Oene GH, et al. Detecting exacerbations using the clinical COPD questionnaire. *Health Qual Life Outcomes*. 2010 [cited 2017 Apr 12];8(1):102. DOI: 10.1186/1477-7525-8-102.
- [15] Global Initiative for Chronic Obstructive Lung Disease (GOLD). Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease. [updated 2015.; cited 2016 May 10]. Available from: <http://goldcopd.org/>.
- [16] Fernandez-Granero MA, Sanchez-Morillo D, Leon-Jimenez A. Computerised analysis of telemonitored respiratory sounds for predicting acute exacerbations of COPD. *Sensors*. 2015;15(10):26978–26996.
- [17] Hadjileontiadis LJ. Lung sounds: an advanced signal processing perspective. California (USA): Morgan&Claypool Publishers; 2008.
- [18] Enderle J, Blanchard S, Bronzino J. Introduction to biomedical engineering. Burlington (USA): Elsevier Academic Press; 2005.
- [19] Hashemi A, Arabalibiek H, Agin K. Classification of wheeze sounds using wavelets and neural networks. *Int Conf Biomed Eng Technol*. 2011;11:127–131.
- [20] Kandaswamy A, Kumar CS, Ramanathan RP, et al. Neural classification of lung sounds using wavelet coefficients. *Comput Biol Med*. 2004;34:523–537.
- [21] Yu L, Liu H. Feature selection for high-dimensional data: a fast correlation-based filter solution. In: Fawcett T, Mishra N, editors. Proceedings of the 20th International Conference on Machine Learning (ICML-03); 2003 Aug 21–24; Washington DC (USA). Menlo Park (CA): AAAI Press; 2003. p. 856–863.
- [22] Bolón-Canedo V, Sánchez-Marño N, Alonso-Betanzos A. Feature selection for high-dimensional data. Berlin, Heidelberg (Germany): Springer; 2015.
- [23] Rokach L. Ensemble methods in supervised learning. In: Maimon O, Rokach L, editors. Data mining and knowledge discovery handbook. Boston (MA): Springer; 2009. p. 959–979.
- [24] Sherrod PH. DTREG—Predictive modeling software-manual. [updated 2014; cited 2016 May 10]. Available from: <https://www.dtreg.com/download/>.
- [25] Anthonisen NR, Manfreda J, Warren CP, et al. Antibiotic therapy in exacerbations of chronic obstructive pulmonary disease. *Ann Intern Med*. 1987;106(2):196–204.
- [26] Fernández-Granero MA, Sánchez-Morillo D, León-Jiménez A, et al. Automatic prediction of chronic obstructive pulmonary disease exacerbations through home telemonitoring of symptoms. *Biomed Mater Eng*. 2014;24:3825–3832.
- [27] Henderson C, Knapp M, Fernández JL, et al. Cost-effectiveness of telecare for people with social care needs: the whole systems demonstrator cluster randomised trial. *Age Ageing*. 2014;43(6):794–800.
- [28] Rixon L, Hirani SP, Cartwright M, et al. A RCT of telehealth for COPD patient's quality of life: the whole system demonstrator evaluation. *Clin Respir J*. 2015;11:2–25.

- [29] Azar AT, El-Metwally SM. Decision tree classifiers for automated medical diagnosis. *Neural Comput Appl*. 2013;23:2387–23403.
- [30] Sharma N, Om H. Data mining models for predicting oral cancer survivability. *Netw Model Anal Health Inform Bioinform*. 2013;2(4):285–295.
- [31] Metz CE. Basic principles of ROC analysis. *Semin Nucl Med*. 1978;8:283–298.
- [32] Seemungal TA, Donaldson GC, Bhowmik A, et al. Time course and recovery of exacerbations in patients with chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*. 2000;161:1608–1613.
- [33] Langsetmo L, Platt RW, Ernst P, et al. Underreporting exacerbation of chronic obstructive pulmonary disease in a longitudinal cohort. *Am J Respir Crit Care Med*. 2008;177:396–401.
- [34] Sanchez-Morillo D, Fernandez-Granero MA, León Jiménez A. Detecting COPD exacerbations early using daily telemonitoring of symptoms and k-means clustering: a pilot study. *Med Biol Eng Comput*. 2015;53:441–451.
- [35] McKinstry B. The use of remote monitoring technologies in managing chronic obstructive pulmonary disease. *QJM Int J Med*. 2013;106(10):883–885.
- [36] Jacome C, Marques A. Computerized respiratory sounds are a reliable marker in subjects with COPD. *Respir Care*. 2015;60(9):1264–1275.
- [37] Jensen MH, Cichosz SL, Dinesen B, et al. Moving prediction of exacerbation in chronic obstructive pulmonary disease for patients in telecare. *J Telemed Telecare*. 2012;18(2):99–103.
- [38] van der Heijden M, Lucas PJF, Lijnse B, et al. An autonomous mobile system for the management of COPD. *J Biomed Inform*. 2013;46(3):458–469.
- [39] 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.
- [40] Mohktar MS, Redmond SJ, Antoniadis NC, 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):51–59.