ORIGINAL ARTICLE

Diagnosis of Respiratory Changes in Cystic Fibrosis Using a Soft Voting Ensemble with Bayesian Networks and Machine Learning Algorithms

Noemi P. Pinto[1] · Jorge L. M. Amaral[2] · Agnaldo J. Lopes[3] · Pedro L. Melo[4]

Abstract

Purpose Advances in the treatment of cystic fibrosis (CF) have allowed patients to reach adulthood. The forced oscillation technique (FOT) is a new method for providing an exam that is simple to perform and simultaneously provides a detailed respiratory system evaluation. The purpose of this study was to use machine learning (ML) algorithms to increase the accuracy and interpretability of FOT parameters in the investigation and diagnosis of respiratory changes in adults with CF.
Methods The database was created based on 150 measurements in 50 volunteers (23 in the control group and 27 in the test group). The following supervised ML algorithms were selected for the tests: K-nearest neighbors (KNN), random forest (RF), AdaBoost with decision trees (ADAB), and light gradient boosting (LGB). These data were also subjected to a Bayesian network synthesized by a genetic algorithm (BNGA) in an attempt to maintain good accuracy and increase the interpretability of the results. A soft vote ensemble strategy was employed to enhance the diagnostic accuracy.
Results The first part of this study showed the best FOT parameter: the reactance X_m (AUC = 0.85), indicating moderate accuracy. In the second part, the original FOT parameters were used as input in the chosen algorithms. BNGA had the best performance alone (AUC = 0.88), while the soft voting ensemble achieved AUC = 0.90. When cross-product and feature selection methods were applied, the RF and BNGA were the algorithms with the best results (AUC = 0.88), and the soft voting ensemble achieved an AUC = 0.94.
Conclusion This study provides high diagnostic accuracy with improved interpretability of the FOT parameters, which assists doctors in the medical diagnosis of respiratory changes in CF.

Keywords Cystic fibrosis · Forced oscillation technique · Respiratory oscillometry · Machine learning · Bayesian networks · Genetic algorithm · Clinical decision support system

1 Introduction

Cystic fibrosis, which is also called mucoviscidosis, is an autosomal recessive genetic disease that affects both men and women and is more common in the Caucasian population. It is caused by mutations in the gene located on chromosome seven, which is responsible for encoding the protein cystic fibrosis transmembrane conductance regulator (CFTR) [1]. This protein is responsible for regulating and participating in the transport of electrolytes through the cellular membranes of the respiratory, digestive, and reproductive systems, although the respiratory system is the most affected [2].

This disease was previously first diagnosed in newborns, and it led to death within the first year of life. Due to advances in the treatment and diagnosis of cystic fibrosis, these patients can now reach adulthood [3].

✉ Pedro L. Melo
plopes@uerj.br

1 Electronic Engineering Post-Graduation Program, State University of Rio de Janeiro, Rio de Janeiro, Brazil

2 Department of Electronics and Telecommunications Engineering, State University of Rio de Janeiro, Rio de Janeiro, Brazil

3 Pulmonary Function Laboratory, Faculty of Medical Sciences, State University of Rio de Janeiro, Rio de Janeiro, Brazil

4 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

Currently, approximately 70,000 adult patients are registered worldwide.

Cystic fibrosis is a progressive disease. Thus, over the years, a patient will present with increased airflow obstruction and respiratory abnormalities. These symptoms contribute to a decreasing life expectancy, causing discomfort during sleep, intolerance to physical activities and even to normal activities in everyday life [1].

The recommended diagnosis of cystic fibrosis is based on three criteria: clinical analysis, the concentration of sodium chloride obtained through the sweat test, and CFTR analysis [4]. Among the diagnostic methods in use, spirometry has also been an essential tool. However, research on new techniques has been a great motivation to improve the detection of cystic fibrosis.

The forced oscillation method, which is also designated as respiratory oscillometry, has been studied to analyze the mechanical properties of the respiratory system [5]. Currently, each of the FOT parameters is used alone to detect respiratory changes, and the attribute that presents the highest performance is selected as a criterion for identifying the disease [6].

The use of machine learning methods associated with oscillometric parameters has brought about significant advances in the diagnosis of respiratory diseases [6–8]. This association, however, has not been investigated in adult patients with cystic fibrosis.

It is also important to note that although oscillometry may provide a simple exam, thereby simplifying patient testing, the interpretation of the oscillometric parameters is difficult, demanding an experienced and trained medical team. This method is so demanding because the results are based on electrical engineering methods, which describe resistance and reactance curves and derivative parameters [5]. For this reason, the interpretation of the result is as vital as the hypothesis given by the model in this problem. This characteristic of expressing the behavior of a system comprehensibly is called interpretability and does not have a performance metric to evaluate [9].

In this context, this work proposes using interpretable machine-learning algorithms to assist medical teams in investigating and diagnosing respiratory changes in patients with CF using the data provided by respiratory oscillometry.

2 Methods

2.1 Research Ethics, Patient Consent and Datasets

The local Medical Research Ethics Committee approved this study, which was developed according to the Declaration of Helsinki.

The biometric parameters, including patient height, weight and age, were obtained from each volunteer at the time of the exams. For inclusion in this study, all the volunteers had to sign informed consent forms.

The dataset used in this work was obtained using a previously described instrument [10]. Oscillometric exams were performed in accordance with international standards [5]. To prevent air leakage and induce normal breathing through the equipment nozzle, individuals were required to wear a nasal clip during the procedure. The exams were performed in 23 individuals in the control group and 27 patients with CF who were part of the test group. For each exam, three measurements were taken, which generated a dataset of 150 instances for the experiments.

2.2 Forced Oscillation Measurements and Parameters

During an FOT exam, the individual should remain seated, use a nasal clip and maintain spontaneous breathing, while a constant flow renews the air inspired by the patient. This method uses small pressure oscillatory signals (less than 2 cmH$_2$O peak-to-peak) that are applied to the respiratory system entrance. The ratio of the Fourier transform (F) of the oscillatory pressure (P) to the oscillatory flow (V') generated from this oscillatory stimulus is used to calculate the input impedance $[\text{Zrs} = F(P)/F(V')]$. Based on this analysis, we can generate resistance and reactance curves as a function of frequency that represent the total mechanical properties of the respiratory system [5].

Resistive respiratory impedance results were interpreted using a linear regression analysis over a range from 4 to 16 Hz. Thus, it is possible to determine the resistance in the intercept at 0 Hz (R_o) and the slope of the linear relationship of resistance versus frequency (S) [11]. These parameters estimate the total resistance and the homogeneity of the respiratory system, respectively [12, 13]. The cited analysis also gives the mean resistance (R_m), which is primarily sensitive to the airway caliber [14].

The interpretation of the reactance curves is made using the mean reactance (X_m) and the resonant frequency (F_r), which are associated with ventilation homogeneity [8] as well as the dynamic compliance (C_{dyn}) and elastance (E_{dyn}). The interpretation also includes the respiratory impedance modulus at 4 Hz ($Z_{4\text{Hz}}$), which is associated with the work of breathing, integrating the resistive and elastic loads in the respiratory system [15].

2.3 Machine Learning Algorithms

Machine learning (ML) is a branch of artificial intelligence that allows computers to learn without being explicitly taught to do so [16]. Its approaches can be used primarily

to address issues with no deterministic solution, with data that are used to allow the algorithms to identify relationships automatically. Previous research has found that using oscillometric features in combination with ML algorithms may be useful in addressing asthma [6], in the differential diagnosis of asthma and restrictive respiratory diseases [7], and in systemic sclerosis [8].

In the present study, the use of ensemble techniques was investigated in addition to the methods used in the aforementioned studies. We wish to investigate light gradient boosting (LGB) [17], a form of ensemble derived from gradient boosting, by emphasizing performance and scalability. Another ensemble strategy employed here is the soft voting ensemble. It trains multiple base models and uses voting to combine the individual predictions to arrive at the final ones. It does not require the base models to be homogenous. In other words, we can train different base learners, for example, a random forest and a K nearest neighbor, and then use the voting ensemble to combine the results. This approach is called the soft voting ensemble because the final class prediction is made based on the average probability calculated using all the base model predictions. Among the studied classifiers, two are chosen to participate in the ensemble. Our strategy consists of selecting classifiers with better performance that are less correlated with the others. We ranked the classifiers in descending order of AUCs and ascending order of the sum of the correlations and chose the two with the smallest sum of the ranks.

The interpretability of a classifier is crucial in research related to respiratory diseases, in addition to producing accurate results. Knowing how classification is performed and how the features interact will help us better understand the diagnosis. Hence, we applied Bayesian networks to capture the relation between the features.

We also evaluated the following algorithms: K-nearest neighbor (KNN), AdaBoost with decision trees (ADAB), random forest (RF), [18] light gradient boosting (LGB) [17] and Bayesian networks [19]. The first three algorithms have already been described previously [6–8, 20]. A concise overview of the two algorithms that have not been employed in earlier studies may be found in the supplement.

The genetic algorithm (GA) is a heuristic technique used to search and optimize complex problems, and it is inspired by Darwin's natural selection theory. The fundamental concept is to create an initial population of individuals that represents potential solutions. These individuals are encoded in chromosomes, which are appraised over generations according to the survival of the fittest concept. Individuals who cannot gain resources via natural selection are unlikely to pass their genes on to future generations. As a result, these people will not leave their offspring. On the other hand, successful individuals have a better chance of passing on their genes to future generations and producing new ones who

have a better chance of surviving. The population of individuals addressed using the GA method reflects the search space, which contains potential solutions. The environment is the problem to be solved, and generations are represented by cycles [21]. All the individuals in the population are evaluated by a fitness function that scores how good a solution is to the problem. For the next generation, the probability of an individual being selected for crossover or mutation operators is calculated by the fitness score. This process is repeated until the stop criterion is reached. Thus, the GA optimizes problems by providing the best solution according to an application, but it does not guarantee the optimal solution. This algorithm can be used with other techniques and applied to various types of problems [22].

2.4 Bayesian Network Synthesized by Genetic Algorithm

The strategy chosen to perform the structure learning of Bayesian networks was the use of genetic algorithms. The joint use of both techniques was implemented and called the BNGA, which aims to create and select the best structure that describes relations among the variables of a problem. The BNGA algorithm generates possible solutions through the random creation of several networks represented through adjacency matrices. These networks are built based on these matrices and have their probability distributions calculated by a BN algorithm. There are primary characteristics that must be defined to use BNGA: chromosome representation, creation of the initial population, fitness function, selection function, and genetic operators.

In BNGA, a chromosome corresponds to the structure of a BN with n variables and to genes formed by a binary code. This structure of a network can be represented by an adjacency matrix of size $n \times n$, in which the elements are described according to the connections between j and i. These existing links between variables ($I \times j = 1$) or nonexisting links ($i \times j = 0$) are expressed in an array that can be decomposed, column by column, to generate a vector [23, 24]. The initial population of the BNGA algorithm is created randomly with a uniform distribution [21]. The fitness function determines how appropriate each generated individual is during the search for the best solution. Each possible solution, as represented by vectors, is received by this function and converted into a sparse matrix. Once the structure is in the matrix format, this algorithm trains and tests the generated structure. Two important pieces of information are provided by this fitness function: the area under the receiver operating characteristic (ROC) curve (AUC) of the tested structure and the score vector with the probability of each sample used during the tests. These probabilities will be used for the construction of an ROC curve.

The selection of individuals is made by the probabilistic roulette method in which the fittest individual has a higher probability of being chosen and forms the next generation. Ranking by geometric normalization was also used to order individuals and prevent the fittest individual from always being chosen, leading the algorithm to premature convergence [25].

Genetic operators are primary search engines used by GA for creating new individuals based on the existing population. One of the main operators is the crossover, which uses two individual parents to generate two new individuals by crossing their chromosomes. For the BNGA algorithm, the simple crossover presented better performance. The mutation operator is also widely used in GA, changing the chromosome of an individual and generating only a new solution for the next generation. Binary mutation was used in BNGA, making changes based on a calculated probability.

2.5 Experimental Design

We conducted our study during five experiments. First, the capability of each FOT parameter to detect respiratory changes in cystic fibrosis correctly was evaluated alone.

In the second experiment, all eight original FOT parameters were applied to ML algorithms to increase the performance. Four of the five chosen classifiers were implemented with Scikit-learn, a machine learning library written in python, and BNGA was implemented in MATLAB with the toolboxes Probabilistic Graphical Model 9.2.3 [26] and GAOT [25]. The measurement of the performance was based on the area under the ROC curve (AUC) because it is one of the most employed metrics in medicine [27] and provides a superior way to compare accuracy of the used classifiers with [28]. Feature selection was not implemented; thus, we used all the FOT parameters. The dataset contains 150 FOT measurements.

Because the dataset contains 150 FOT measurements, the k-fold validation procedure [29] is adequate for evaluating the generalization proficiency in the whole dataset.

An important step in model selection is hyperparameter tuning. For this purpose, Scikit-learn possesses several strategies, such as grid search, which tests all possible hyperparameter associations. Table 1 describes the classifiers and their respective hyperparameters used for tuning.

For the third experiment, a smaller set was selected from the original FOT parameters, aiming for better algorithms performance. This technique was performed using the wrapper strategy, which provides input parameters that optimize the average AUC. The search for this set can demand high computation costs. Consequently, many strategies are applied to this effort and for feature selection. This process can also cause overfitting. Therefore, cross-validation was also used during this experiment. The feature selection

Table 1 Hyperparameter values for tuning

Classifier	Hyperparameters chosen for tuning	Hyperparameter values
KNN	Number of neighbors	3, 5, 7, 9,15
AdaBoost	Number of base estimators	10, 30, 60, 100, 200, 400
	Max depth of base estimators	3, 4, 5, 10, 15
RF	Number of base estimators	10, 30, 60, 100, 200, 400
	Max depth of base estimators	3, 4, 5, 10, 15
LGB	Number of base estimators	10, 30, 60, 100, 200, 400
	Max depth of base estimators	3, 4, 5, 10, 15
	Subsample	0.6, 0.8, 1.0
	Colsample_bytree	0.6, 0.8, 1.0

procedure was performed in each classifier during the training, which used tenfold cross-validation. The training was repeated ten times, by selecting one folder for the test and the other folders for the training set. Internal cross-validation, which uses only the training set, was applied to select the best parameters for each classifier. This process was used in each test folder.

In the fourth and fifth experiments, the input feature set was the cross-product of the input parameters used in the second and third experiments. Through this method, the classifiers would result in improved performance.

During the first experiment, the best FOT parameter (BFP) performance was selected for comparison with the five other classifiers (K-NN, RF, AdaBoost, LGB and BNGA) of the second, third, fourth, and fifth experiments. In the clinical scenario, the severity of respiratory diseases, such as chronic obstructive pulmonary disease (COPD) [30], is currently classified using one feature, motivating this choice. MedCalc 8.2 software (Medicalc Software, Mariakerke, Belgium) was used to compare the AUC values obtained during the experiments through the methodology described in Delong et al. [31].

3 Results

There were no significant biometric differences among the groups (Table 2). As expected, the spirometric parameters decreased in patients with CF ($p < 0.04$).

3.1 Forced Oscillation Parameters

The bar charts in Fig. 1 describe the oscillometric results of the control and test groups. The mean values of each FOT parameter were calculated at a 95% confidence interval. Using analysis of variance (ANOVA), all the parameters of the FOT showed a significant difference in comparison with the test group ($p < 0.001$). The mean values of R_o, R_m, Z_{4Hz},

Table 2 Biometric and spirometric parameters of the studied groups

	Control group	Cystic fibrosis	p
Age (years)	25.6 ± 3.1	25.0 ± 5.7	ns
Weight (kg)	64.4 ± 12.8	59.4 ± 13.7	ns
Height (cm)	167.1 ± 8.9	166.8 ± 8.9	ns
Body mass index	23.0 ± 2.8	21.2 ± 3.8	ns
Male/female	14/9	16/11	–
FEV_1 (L)	4.2 ± 1.0	2.3 ± 1.1	<0.0001
FEV_1 (%)	101.7 ± 13.0	65.8 ± 28.6	<0.0001
FVC (L)	3.7 ± 0.8	3.5 ± 1.3	0.03297
FVC (%)	105.2 ± 13.8	84.1 ± 25.4	0.00317
FEV_1/FVC (%)	88.7 ± 4.4	64.4 ± 13.9	<0.0001
FEF/FVC (%)	103.3 ± 5.1	75.0 ± 16.3	<0.0001

FEV_1 (L): forced expiratory volume during the first second (liters);

FEV_1 (%): forced expiratory volume during the first second (percentile of the predicted values);

FEV_1/FVC (%): forced expiratory volume during the first second divided by the forced vital capacity (percentile values);

FEF/FVC (%): forced expiratory flow between 25 and 75% divided by the forced vital capacity (percentile values); Values are presented as the means ± standard deviation

F_r and E_{dyn} increased in the test group compared to the control group. Therefore, we can suppose that individuals with cystic fibrosis usually have higher values of resistance (R_o and R_m), impedance (Z_{4Hz}), resonance frequency (F_r), and elastance (E_{dyn}) than the controls. However, the mean values of X_m, C_{dyn}, and S from the test group decreased compared to the control group. In this case, we can suppose that patients have more negative values for reactance (X_m) and resistance curve slope (S) and lower values for dynamic compliance (C_{dyn}).

3.2 First Experiment: Diagnostic Accuracy of Each FOT Parameter

The values obtained in this first experiment are summarized in Fig. 2. All the parameters presented moderate diagnostic accuracy ($0.70 \leq AUC \leq 0.90$). X_m and F_r presented the best performance, with AUC values = 0.85 and 0.84, respectively. The ROC curves of each FOT parameter, the AUC with the standard error, the confidence intervals, the sensitivity and the specificity can be found in the supplementary material (Fig. S1).

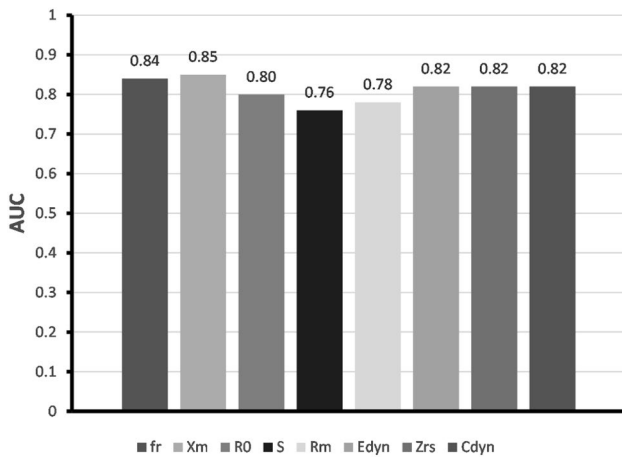**Fig. 1** Comparison of FOT parameters from the control group (CG) and the test group (TG)
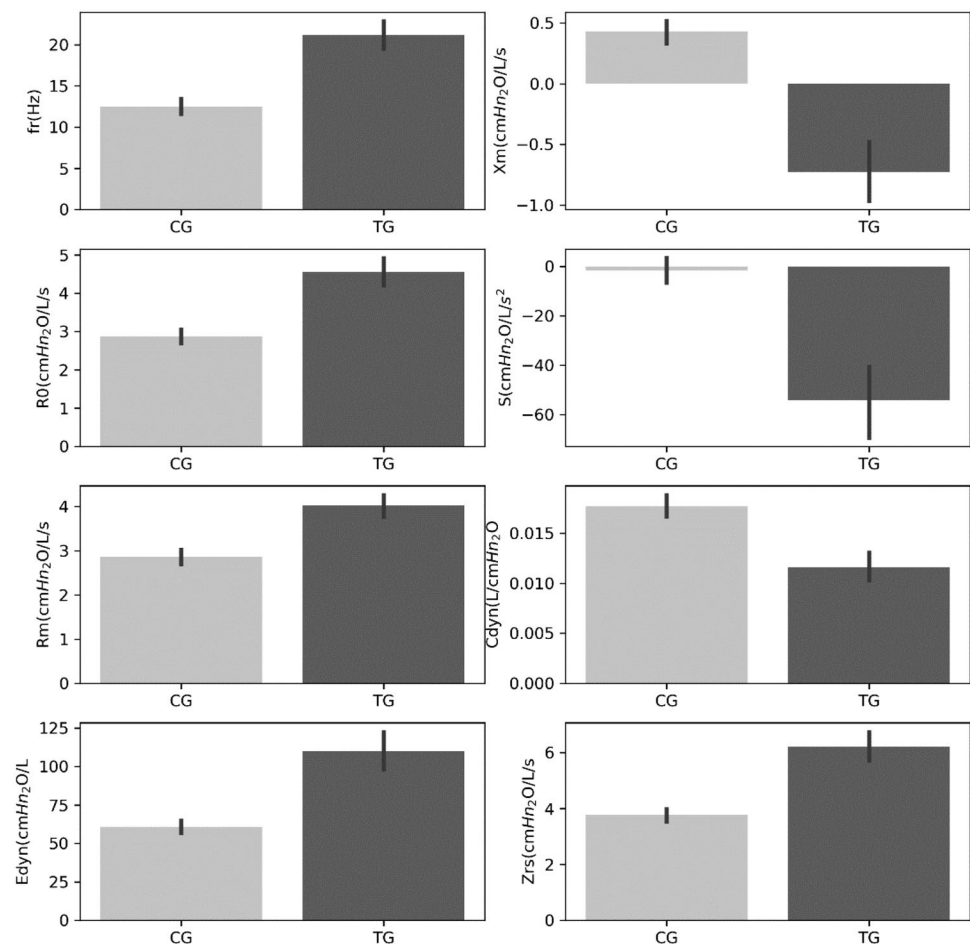
Fig. 2 Experiment 1—AUC for each of the FOT parameters

3.3 Second Experiment: Effect of Machine Learning Methods on Diagnostic Accuracy

The average ROC curves of the BFP (X_m) and the best classifiers obtained in this experiment are shown in Fig. 3. Among the individual algorithms, BNGA presented the best performance, with an AUC equal to 0.88. ADAB and BNGA had the lowest sum of ranks, and they were chosen to compose the soft voting ensemble (ENSEMBLE) that achieved an AUC = 0.9. More details about the ranks are provided in the supplementary material (Fig. S2).

3.4 Third Experiment: Effect of Machine Learning Methods Associated with Feature Selection on Diagnostic Accuracy

Figure 4 shows the AUCs for the BFP (X_m) and the studied classifiers (K-NN, ADAB, RF, LGB and BNGA) with feature selection. KNN has the best performance (AUC = 0.86). A soft voting ensemble (ENSEMBLE) was composed of the KNN and BNGA and achieved an AUC = 0.9. An ROC curve comparison showed a statistically significant difference between the BFP and ENSEMBLE, with a p value < 0.05. More details of this analysis may be obtained in the supplementary material (Fig. S3).

3.5 Fourth Experiment: Effect of the Cross Products and Machine Learning Methods on Diagnostic Accuracy

Thirty-six combinations of the cross products were generated for this experiment. To represent a possible solution in the BNGA algorithm, 37×37 matrices were needed. During the marginalization of the network, the junction tree method [32], which is provided by the PGM toolbox, performs several processes that require a high computational cost. Therefore, the BNGA algorithm did not converge. However, there were no failures, and the experiment could be performed using the other algorithms.

The AUCs of the BFP and the classifiers studied are shown in Fig. 5. Using the cross products as an input, only the K-NN performed slightly better (AUC = 0.86) than the BFP. In addition, a soft voting ensemble (ENSEMBLE) was composed of the KNN and LGB and achieved an
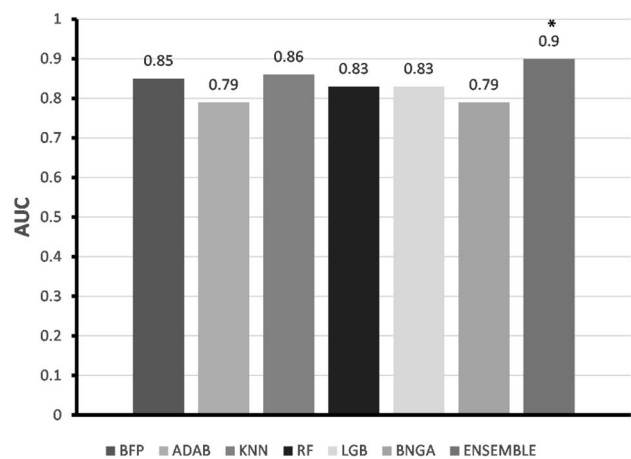

Fig. 3 Experiment 2—Diagnostic accuracy of all eight FOT parameters associated with machine learning techniques. AUCs for the best FOT parameter (BFP) and for the ML methods. Diagnostic accuracy of all eight FOT parameters associated with machine learning techniques

Fig. 4 Experiment 3—Diagnostic accuracy of the best original FOT parameters selected by recursive feature selection associated with machine learning techniques. AUCs for the best FOT parameter (BFP) and for the ML methods. Additionally, "*" indicates that there is a statistically significant difference in relation to the BFP ($p < 0.05$)
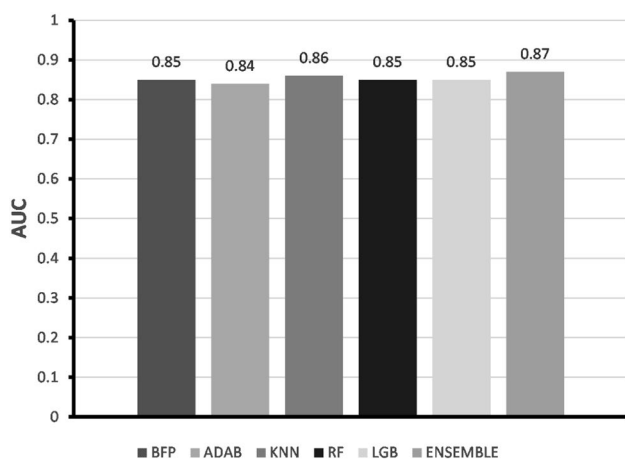
Fig. 5 Experiment 4—Diagnostic accuracy of the cross products of the eight FOT parameters associated with machine learning techniques. AUCs for the best FOT parameter (BFP) and for the ML methods
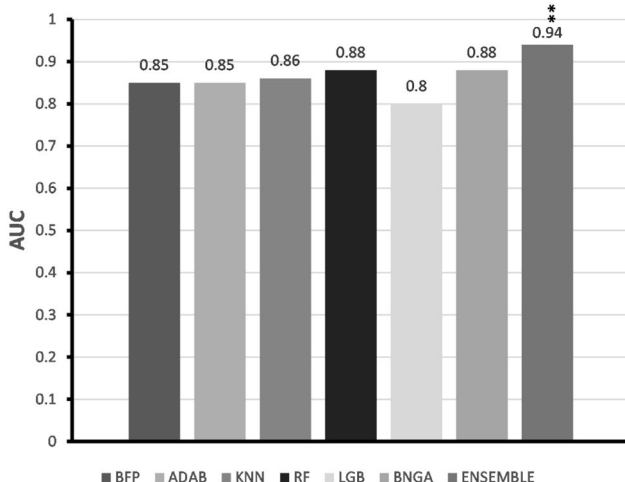

Fig. 6 Experiment 5—Diagnostic accuracy of cross products from original FOT parameters selected by recursive feature selections associated with machine learning techniques. AUCs for the best FOT parameter (BFP) and for the ML methods. Additionally, "**" indicates that there is a statistically significant difference in relation to the BFP ($p < 0.01$)

AUC = 0.87. Detailed descriptions of the ROC curves are presented in the supplement (Fig. S4).

3.6 Fifth Experiment: Effect of the Cross Products from the Best Parameters in Association with Machine Learning on Diagnostic Accuracy

Figure 6 presents the AUC of the BFP and of the evaluated algorithms with feature selection in the cross products of the FOT parameters.

Regarding the individual classifiers, BNGA and RF obtained the best results (AUC = 0.88 and AUC = 0.87). Remarkably, the ENSEMBLE, which combines RF and BNGA, achieved an AUC = 0.94. The statistical test showed that there was a statistically significant difference between BFP and ENSEMBLE, with a p value < 0.01. A detailed description of the resulting ROC curves is presented in the supplement (Fig. S5).

Figures 7 and 8 show Se at a moderate Sp (Sp = 75%) and Se at a higher Sp (Sp = 90%), respectively.

4 Discussion

Machine learning methods have a long history of contributing to lung function analysis [20]. The present study expands this contribution by developing clinical decision support systems to improve the diagnostic accuracy and simplify the clinical use of FOT in cystic fibrosis. During the experiments, the KNN, ADAB, and BNGA classifiers presented AUC values higher than those obtained by the best FOT parameter, achieving a high diagnostic accuracy. In addition, the soft voting ensemble (ENSEMBLE) achieved superior performance in all experiments.

The respiratory changes observed in CF patients (Fig. 1, Table 2) were consistent with the underlying physiology [2, 3]. The first experiment showed respiratory reactance (X_m) as the FOT parameter that presented the highest accuracy (Fig. 2, AUC = 0.85).

In the second experiment (Fig. 3), we used all the parameters provided by the FOT as attributes. The best individual result was presented by the BNGA algorithm (AUC = 0.88), and the ENSEMBLE obtained AUC = 0.90.

During the third experiment (Fig. 4), the best FOT parameters were used as input in all the classifiers, and they coincided with the feature selection made by a specialist. Altogether, five parameters were selected: R_o, R_m, X_m, C_{din} and Z_{4Hz}. KNN was the algorithm with the best performance (AUC = 0.86), but the BNGA algorithm showed the lowest performance (AUC = 0.79). The ENSEMBLE presented AUC = 0.90, achieving a statistically significant increase in comparison with the BFP.

As an attempt to improve the performance of algorithms, the cross-product of original FOT parameters was used in the fourth experiment (Fig. 5), providing a dataset in a higher dimension with 36 combinations generated by this method. The KNN classifier presented the best performance (AUC = 0.86), and ENSEMBLE attained an AUC = 0.87. The BNGA algorithm could not converge during this experiment because of the computational effort necessary for the network marginalization process used by the junction tree algorithm. This limitation can also be observed in other works using Bayesian networks, as in the article by Silander

Fig. 7 Summary of the experiments describing sensitivity comparisons at 75% Sp obtained using the best FOT parameter (BFP) and ML methods in all experiments
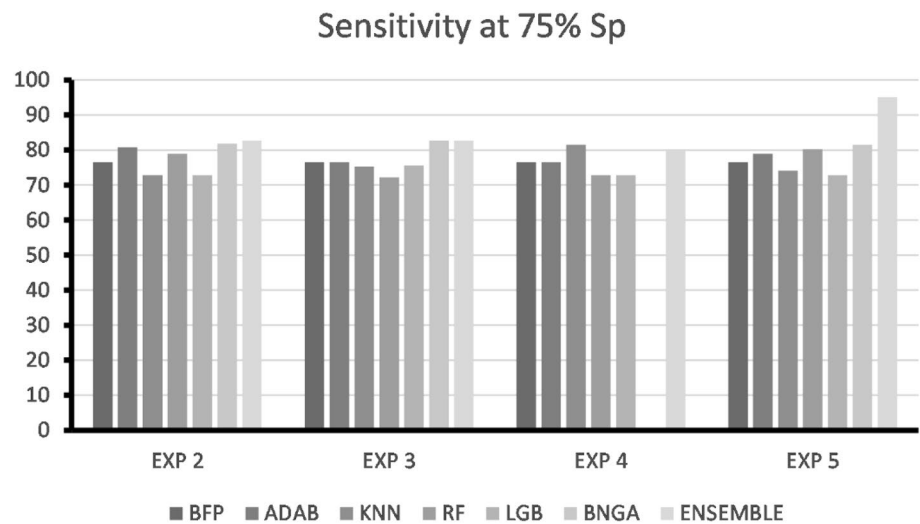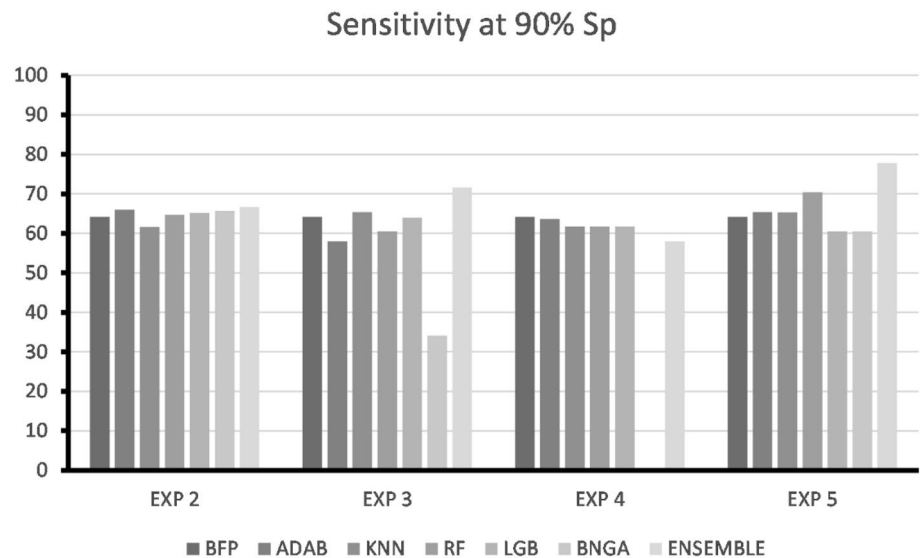

Fig. 8 Summary of the experiments describing comparisons of the sensitivity at 90% Sp obtained using the best FOT parameter (BFP) and ML methods in all experiments

and Myllymaki [33], in which the maximum number supported by the model is 30 features.

During the fifth experiment (Fig. 6), the use of the cross-product method in the best FOT parameters of the third experiment generated 15 combinations for the input of the classifiers. The RF and BNGA algorithms had the best results, presenting AUC values of 0.88. ENSEMBLE presented an AUC = 0.94, and the comparison of the ROC curves between BFP and ENSEMBLE showed a significant improvement ($p < 0.01$).

As shown in Figs. 7 and 8, at least two algorithms reached the range of moderate Se (70 to 90%) in the second and fifth experiments with the best results. In both cases, Se and Sp obtained better results when compared to the best individual FOT parameter. At least one algorithm reached the range of moderate Se in the third and fourth experiments. In all the experiments, ENSEMBLE presented Se values greater than

or equal to those of the individual algorithms, and in the fifth experiment, it achieved Se > 90%.

The soft voting ensemble achieved high diagnostic accuracy (AUC ≥ 0.9) in three experiments, which indicates that the strategy of combining classifiers with higher AUCs that were less correlated with the others was successful. In addition, we showed that BNGA was less correlated with the other machine learning algorithms, and therefore, it helped to introduce diversity to the soft voting ensemble. This finding indicates that it provided important information when the other algorithms did not.

The main disadvantage of the BNGA is the time required to compute the Bayesian networks with the help of genetic algorithms (GA). As mentioned before, the marginalization of the network, the junction tree algorithm provided by the PGM toolbox, performs several processes requiring a high computational cost. Its worst-case complexity is exponential: $O(a^c n^b)$,

Table 3 Cutoff points for the discretization of FOT parameters, means and standard deviation

Parameter	Cutoff point	Mean ($\pm$ standard deviation)
R_o	3.31	3.78 ± 1.54
R_m	3.21	3.48 ± 1.13
X_m	0.18	-0.19 ± 1.04
C_{dyn}	0.014	0.014 ± 0.007
S	-10.25	-30.15 ± 57.61
Z_{4Hz}	4.44	5.10 ± 2.25
F_r	14.19	17.20 ± 7.50
E_{dyn}	72.54	87.40 ± 48.80

where a and b are constants, n is the number of attributes, and c is the largest clique of the junction tree. In addition to this complexity, GA requires the junction tree algorithm to be executed several times. Suppose the number of generations is indicated by g and the number of individuals in the population is p. The number of folds is k in the k-fold cross-validation. In that case, the total complexity of BNGA is $O(gpka^cn^b)$. That is why the BNGA took up to 2 h and 33 min in the second experiment, considering that $g = 20$ and $p = 15$ are fairly modest numbers for a GA experiment. To provide a comparison, the time it took to search for hyperparameters and train all the other classifiers together was 2 h and 9 min. Nevertheless, the Bayesian network synthesized by BNGA provided a crucial diversity that allowed the ensemble to reach higher AUCs.

In addition to the AUC values, the interpretability could also be analyzed through the Bayesian networks constructed and selected by genetic algorithms. Even when trained with a limited dataset, the BNGA algorithm proved its efficiency, presenting conditional probabilities that can describe the characteristics of the respiratory system of an individual with cystic fibrosis.

The use of FOT parameters in Bayesian networks requires that all instances must be discretized. Table 3 shows the cutoff points. The dataset was labeled as follows: values below the respective cutoff point were labeled as 1, representing lower values that the variable can assume. The values above the respective cutoff point were labeled as 2, representing the highest values of the variable. For the class, the control group was labeled as 0, and the test group was labeled as 1. Based on this information, the discrete FOT parameters can be summarized according to Table 4.
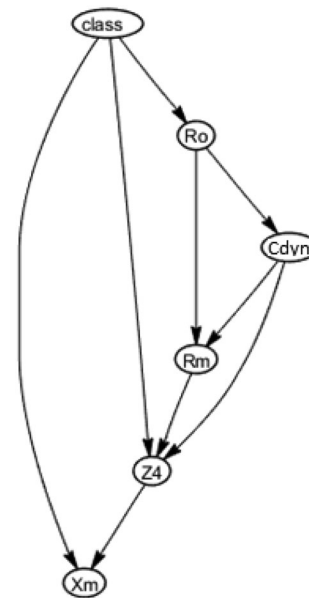

Fig. 9 Structure constructed with the best FOT parameters

Table 5 A priori probabilities of class variables in the best FOT parameter network

P(class = 0)	P(class = 1)
0.49	0.51

A graphical analysis of the relationship among FOT parameters can be performed through the networks provided by the BNGA algorithm. This network was selected for analysis based on the minimum number of arcs among variables. This choice makes the visual inference simpler and the joint probability distribution tables (JPD) smaller. In this analysis, the chosen structure was generated during the third experiment using the best FOT parameters (Fig. 9). This network has six JPD tables, in which the possible biomechanical combinations are highlighted.

Table 5 shows the a priori probabilities of the class node, in which the probability of an individual not suffering from cystic fibrosis is 0.49 and the probability of being a patient is 0.51. Tables 6, 7, 8, 9 and 10 present the JPD calculated for the best FOT parameter nodes.

Let us calculate the probability for the general behavior in test group P ($class = 0$, $R0 = 1$, $Rm = 1$, $Cdyn = 2$, $Z4 = 1$, $Xm = 2$):

Table 4 General behavior of characteristics in the control and test groups

	R_o	R_m	Z_{4Hz}	F_r	E_{dyn}	X_m	C_{dyn}	S	$Class$
Control group	1	1	1	1	1	2	2	2	0
Test group	2	2	2	2	2	1	1	1	1

Table 6 JPD for the R_o variable in the best FOT network

	P($R_o = 1$|class)	P($R_o = 2$|class)
Class = 0	0.94	0.06
Class = 1	0.47	0.53

Table 7 JPD for the C_{dyn} variable in the best FOT network

	P($C_{dyn} = 1$|R_o)	P($C_{dyn} = 2$|R_o)
$R_o = 1$	0.13	0.87
$R_o = 2$	0.84	0.16

Table 8 JPD for the R_m variable in the best FOT network

	P($R_m = 1$|C_{dyn}, R_o)	P($R_m = 2$|C_{dyn}, R_o)
$C_{dyn} = 2$, $R_o = 1$	0.98	0.02
$C_{dyn} = 1$, $R_o = 2$	0.18	0.82

Table 9 JPD for the X_m variable in the best FOT network

	P($X_m = 1$|Z_{4Hz}, class)	P($X_m = 2$|Z_{4Hz}, class)
$Z_{4Hz} = 1$, class = 0	0.02	0.98
$Z_{4Hz} = 2$, class = 1	0.80	0.20

Table 10 JPD for the Z_{4Hz} variable in the best FOT network

	P($Z_{4Hz} = 1$|R_m, C_{dyn}, class)	P($Z_{4Hz} = 2$|R_m, C_{dyn}, class)
$R_m = 1$, $C_{dyn} = 2$, class = 0	0.99	0.01
$R_m = 2$, $C_{dyn} = 1$, class = 1	0.03	0.97

$$P(class = 0, R0 = 1, Rm = 1, Cdyn = 2, Z4 = 1, Xm = 2) =$$

$$P(Xm = 2|Z4 = 1, class = 0) \cdot P(Z4 = 1|Rm = 1, Cdyn = 2)$$

$$\cdot P(Rm = 1|Cdyn = 2, R0 = 1).$$

$$P(Cydn = 2|R0 = 1) \cdot P(R0 = 1|class = 0) \cdot P(class = 0)$$

Using the given tables, P(class = 0, R0 = 1, Rm = 1, Cdyn = 2, Z4 = 1, Xm = 2) = (0.98). (0.99). (0.98).(0.87). (0.94). (0.49) = 0.38.

If one changes one of the FOT parameters, for example, R0 to 2, then P(class = 0, R0 = 2, Rm = 1, Cdyn = 2, Z4 = 1, Xm = 2) would be:

$$P(class = 0, R0 = 2, Rm = 1, Cdyn = 2, Z4 = 1, Xm = 2) =$$

$$P(Xm = 2|Z4 = 1, class = 0) \cdot P(Z4 = 1|Rm = 1, Cdyn = 2)$$

$$\cdot P(Rm = 1|Cdyn = 2, R0 = 2).$$

$$P(Cydn = 2|R0 = 2) \cdot P(R0 = 2|class = 0) \cdot P(class = 0)$$

Using the given tables, $P(class = 0, R0 = 1, Rm = 1, Cdyn = 2, Z4 = 1, Xm = 2) = (0.98). (0.99). (0.18). (0.16). (0.06). (0.49) = 0.00082$.

This result indicates that this combination of FOT parameters is highly unlikely to be observed. Hence, it can help in the reasoning regarding the value of the FOT parameters.

One of the main limitations to the wide clinical use of FOT is the interpretation of its indices, which requires training and experience of the medical team. The present work showed that using Bayesian networks provides interpretability to the result, showing the existing relationships among variables that describe the biomechanics of the respiratory system. Through the generated structures, it is possible to quantify and understand how these variables are related, still maintaining good accuracy in the detection of respiratory changes in patients with cystic fibrosis. Thus, new information is generated, and, in addition to current methods, it can be used to assist medical staff in the study of cystic fibrosis patients, thus simplifying the use of FOT.

5 Conclusions

In summary, five machine-learning algorithms were evaluated to improve the medical services, assisting in the diagnosis of respiratory changes in cystic fibrosis. The individual use of FOT parameters is not efficient for the accurate diagnosis of patients. The use of KNN, RF, and BNGA classifiers allowed us to increase the accuracy, almost reaching the high diagnostic accuracy range in the clinical diagnosis of cystic fibrosis. In addition to the accuracy, the BNGA algorithm provides a helpful network that shows the relationships and the conditional probabilities among FOT parameters. This information may explain the respiratory changes of an individual and may simplify the use of FOT. The soft voting strategy was capable of achieving a high diagnostic accuracy range (AUC ≥ 0.9).

6 Next Steps of the Research

Future studies include (1) the use of another method for the network marginalization process, which requires lower computational effort, (2) in addition to the genetic algorithm, applying other metaheuristics for the creation and selection of structures of Bayesian networks, (3) the implementation of the BNGA classifier in Python and (4) developing an online platform for other researchers to submit their datasets and obtain their models.

Declarations

Conflict of interest The authors have no conflicts of interest to declare.

Ethical Approval The local Medical Research Ethics Committee approved this study, which was developed according to the Declaration of Helsinki.

Informed Consent For inclusion in this study, all the volunteers had to sign informed consent forms.

References

1. Castellani, C., Cuppens, H., Macek, M., Cassiman, J. J., Kerem, E., Durie, P., Tullis, E., Assael, B. M., Bombieri, C., Brown, A., et al. (2008). Consensus on the use and interpretation of cystic fibrosis mutation analysis in clinical practice. *Journal of Cystic Fibrosis, 7*(3), 179–196.
2. Hodson, M. E., Geddes, D. M., & Bush, A. (2007). *Cystic fibrosis* (3rd ed.). Hodder Arnold.
3. Lima, A. N., Faria, A. C. D., Lopes, A. J., Jansen, J. M., & Melo, P. L. (2015). Forced oscillations and respiratory system modeling in adults with cystic fibrosis. *BioMedical Engineering OnLine, 14*(1), 1–18.
4. Farrell, P. M., White, T. B., Ren, C. L., Hempstead, S. E., Accurso, F., Derichs, N., Howenstine, M., McColley, S. A., Rock, M., Rosenfeld, M., et al. (2017). Diagnosis of cystic fibrosis: Consensus guidelines from the cystic fibrosis foundation. *The Journal of Pediatrics, 181*, S4-S15.e11.
5. King, G. G., Bates, J., Berger, K. I., Calverley, P., de Melo, P. L., Dellaca, R. L., Farre, R., Hall, G. L., Ioan, I., Irvin, C. G., et al. (2020). Technical standards for respiratory oscillometry. *The European Respiratory Journal, 55*(2), 1900753.
6. Amaral, J. L. M., Lopes, A. J., Veiga, J., Faria, A. C. D., & Melo, P. L. (2017). High-accuracy detection of airway obstruction in asthma using machine learning algorithms and forced oscillation measurements. *Computer Methods and Programs in Biomedicine, 144*, 113–125.
7. Amaral, J. L. M., Sancho, A. G., Faria, A. C. D., Lopes, A. J., & Melo, P. L. (2020). Differential diagnosis of asthma and restrictive respiratory diseases by combining forced oscillation measurements, machine learning and neuro-fuzzy classifiers. *Medical & Biological Engineering & Computing, 58*(10), 2455–2473.
8. Andrade, D. S. M., Ribeiro, L. M., Lopes, A. J., Amaral, J. L. M., & Melo, P. L. (2021). Machine learning associated with respiratory oscillometry: A computer-aided diagnosis system for the detection of respiratory abnormalities in systemic sclerosis. *BioMedical Engineering OnLine, 20*(1), 31.
9. Carvalho, D. V., Pereira, E. M., & Cardoso, J. S. (2019). Machine learning interpretability: A survey on methods and metrics. *Electronics, 8*(8), 832.
10. de Melo, P. L., Werneck, M. M., & Giannella-Neto, A. (2000). New impedance spectrometer for scientific and clinical studies of the respiratory system. *Review of Scientific Instruments, 71*(7), 2867–2872.
11. Lima, A. N., Faria, A. C., Lopes, A. J., Jansen, J. M., & Melo, P. L. (2015). Forced oscillations and respiratory system modeling in adults with cystic fibrosis. *Biomedical Engineering Online, 14*, 11.
12. Lorino, A. M., Zerah, F., Mariette, C., Harf, A., & Lorino, H. (1997). Respiratory resistive impedance in obstructive patients: Linear regression analysis vs viscoelastic modelling. *The European Respiratory Journal, 10*(1), 150–155.
13. Peslin, R., Hannhart, B., & Pino, J. (1981). Mechanical impedance of the chest in smokers and non-smokers. *Clinical Research, 17*(1), 93–105.
14. MacLeod, D., & Birch, M. (2001). Respiratory input impedance measurement: Forced oscillation methods. *Medical & Biological Engineering & Computing, 39*(5), 505–516.
15. Kaczka, D. W., & Dellaca, R. L. (2011). Oscillation mechanics of the respiratory system: Applications to lung disease. *Critical Reviews in Biomedical Engineering, 39*(4), 337–359.
16. Michalski, R. S., Carbonell, J. G., & Mitchell, T. M. (2013). *Machine learning: An artificial intelligence approach*. Springer.
17. Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems, 30*, 3146–3154.
18. Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning*. Springer.
19. Pourret, O., Na, P., & Marcot, B. (2008). *Bayesian networks: A practical guide to applications*. Wiley.
20. Amaral, J. L. M., & Melo, P. L. (2020). Clinical decision support systems to improve the diagnosis and management of respiratory diseases. In D. Barh (Ed.), *Artificial intelligence in precision health*. Elsevier.
21. Katoch, S., Chauhan, S. S., & Kumar, V. (2021). A review on genetic algorithm: Past, present, and future. *Multimedia Tools and Applications, 80*(5), 8091–8126.
22. Haldurai, L., Madhubala, T., & Rajalakshmi, R. (2016). A study on genetic algorithm and its applications. *International Journal of Computer Sciences and Engineering, 4*(10), 139.
23. Larranaga, P., Poza, M., Yurramendi, Y., Murga, R. H., & Kuijpers, C. M. H. (1996). Structure learning of Bayesian networks by genetic algorithms: A performance analysis of control parameters. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 18*(9), 912–926.

24. Vafaee, F. (2014). Learning the structure of large-scale Bayesian networks using genetic algorithm. In: *Proceedings of the 2014 annual conference on genetic and evolutionary computation* (pp. 855–862).

25. Houck, C. R., Joines, J., & Kay, M. G. (1995). A genetic algorithm for function optimization: A Matlab implementation. *Ncsu-ie tr, 95*(09), 1–10.

26. Armen, A. (2011). Mens X Machina: Probabilistic graph model toolbox, version 0.9.2.3.

27. Obuchowski, N. A., & Bullen, J. A. (2018). Receiver operating characteristic (ROC) curves: Review of methods with applications in diagnostic medicine. *Physics in Medicine & Biology, 63*(7), 07TR01.

28. Japkowicz, N., & Shah, M. (2011). *Evaluating learning algorithms: A classification perspective*. Cambridge University Press.

29. Wong, T.-T., & Yeh, P.-Y. (2020). Reliable accuracy estimates from *k*-fold cross validation. *IEEE Transactions on Knowledge and Data Engineering, 32*(8), 1586–1594.

30. Vestbo, J., Hurd, S. S., Agustí, A. G., Jones, P. W., Vogelmeier, C., Anzueto, A., Barnes, P. J., Fabbri, L. M., Martinez, F. J., & Nishimura, M. (2013). others: Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. *American Journal of Respiratory and Critical Care Medicine, 187*(4), 347–365.

31. DeLong, E. R., DeLong, D. M., & Clarke-Pearson, D. L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics, 44*, 837–845.

32. Barber, D. (2004). Probabilistic modelling and reasoning: The junction tree algorithm. *Course notes*.

33. Silander, T., & Myllymaki, P. (2012). A simple approach for finding the globally optimal Bayesian network structure. https://arxiv.org/abs/1206.6875.