

Machine Learning Associated With Respiratory Oscillometry: A Computer-Aided Diagnosis System for the Detection of Respiratory Abnormalities in Systemic Sclerosis

Domingos Andrade

Rio de Janeiro State University: Universidade do Estado do Rio de Janeiro

Luigi Ribeiro

Rio de Janeiro State University: Universidade do Estado do Rio de Janeiro

Agnaldo Lopes

Rio de Janeiro State University: Universidade do Estado do Rio de Janeiro

Jorge Amaral

Rio de Janeiro State University: Universidade do Estado do Rio de Janeiro

Pedro Lopes de Melo (✉ plopeslib@gmail.com)

Universidade do Estado do Rio de Janeiro <https://orcid.org/0000-0002-2081-6619>

Research

Keywords: System identification techniques, clinical decision support system, machine learning, forced oscillation technique, diagnostic of respiratory diseases, systemic sclerosis, respiratory oscillometry

Posted Date: January 15th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-144194/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published on March 25th, 2021. See the published version at <https://doi.org/10.1186/s12938-021-00865-9>.

Abstract

Background

The use of machine learning (ML) methods would improve the diagnosis of respiratory changes in systemic sclerosis (SSc). This paper evaluates the performance of several ML algorithms associated with the respiratory oscillometry analysis to aid in the diagnostic of respiratory changes in SSc. We also find out the best configuration for this task.

Methods

Oscillometric and spirometric exams were performed in 82 individuals, including controls (n=30), and patients with systemic sclerosis with normal (n=22) and abnormal (n=30) spirometry. Multiple instance classifiers and different supervised machine learning techniques were investigated, including k-nearest neighbours (KNN), random forests (RF), AdaBoost with decision trees (ADAB), and Extreme Gradient Boosting (XGB).

Results and discussion

The first experiment of this study showed that the best oscillometric parameter (BOP) was dynamic compliance. In the scenario Control Group versus Patients with Sclerosis and normal spirometry (CGvsPSNS), it provided moderate accuracy (AUC=0.77). In the scenario Control Group versus Patients with Sclerosis and Altered spirometry (CGvsPSAS), the BOP obtained high accuracy (AUC=0.94). In the second experiment, the ML techniques were used. In CGvsPSNS, KNN achieved the best result (AUC=0.90), significantly improving the accuracy in comparison with the BOP ($p<0.01$), while in CGvsPSAS, RF obtained the best results (AUC=0.97), also significantly improving the diagnostic accuracy ($p<0.05$). In the third, fourth, fifth, and sixth experiments, the use of different feature selection techniques allowed us to spot the best oscillometric parameters. They all show a small increase in diagnostic accuracy in CGvsPSNS, respectively 0.87, 0.86, 0.82, 0.84, while in the CGvsPSAS, the performance of the best classifier remained the same (AUC=0.97).

Conclusions

Oscillometric principles combined with machine learning algorithm provides a new method for the diagnosis of respiratory changes in patients with systemic sclerosis. The findings of the present study provide evidence that this combination may play an important role in the early diagnosis of respiratory changes in these patients.

Background

Systemic sclerosis (SSc) is a chronic connective tissue disease, characterized by thickening and fibrosis of the skin and internal organs such as heart, lungs, kidneys and gastrointestinal tract [1, 2]. Pulmonary complications are the most common causes of death in SSc, and pulmonary arterial hypertension has

become the most crucial life-threatening complication of the disease. The most common pulmonary manifestation is interstitial lung disease, associated with pulmonary fibrosis, where the lungs lose its compliance. This abnormality occurs in approximately 80% of cases, and is associated with reduced survival [2].

The Forced Oscillation Technique (FOT), also known as Respiratory Oscillometry, is a system identification method used to evaluate the respiratory system resistance and reactance. This method provides a detailed analysis of the mechanical properties of the respiratory system, addressing different properties from that evaluated by spirometry, the most traditional method of analysing respiratory diseases. Indeed, oscillometry is likely complementary to spirometry [3]. The measurement is based on the application of low-pressure oscillations to the airway opening to stimulate the respiratory system and measuring the associated flow response. Therefore, this technique requires minimal cooperation and no forced expiratory manoeuvres and can be used in situations when standard measurements of lung function by spirometry is difficult or not feasible, including children, the elderly and patients in advanced stages of disease [4].

Even in the presence of these important clinical advantages, oscillometry is not currently widely used in pulmonary function tests. This limitation arises because this method is based on concepts derived from the Electrical Engineering area, which are not easily interpreted in the clinical environment. Thus, although oscillometry exams are simple, the interpretation of resistance and reactance curves and the derived parameters is a difficult task for the busy untrained pulmonologist, requiring training and experience.

Machine learning (ML) algorithms has been offered an important contribution to the improvement of the lung function tests [5]. In the particular case of oscillometry, previous studies provided clear evidence that these algorithms contributes to simplify the interpretation of the results, and therefore, the clinical use of oscillometry [6, 7]. There is also evidence that the use of these algorithms may help to increase diagnostic accuracy [6, 8]. In spite of the high potential of the combination of these two methods in the study of the lung diseases, there is no previous studies using oscillometry in a combination with ML methods to diagnose respiratory changes in SSc.

In this context, we hypothesized that the use of ML methods associated with respiratory oscillometry analysis would improve the diagnosis of respiratory changes in systemic sclerosis. This paper has two key aims. First, to evaluate several ML algorithms to aid in the diagnostic of respiratory changes in SSc. Second, to find out the best configuration for this task.

In the next four sections of this paper, it is initially provided a description of the patient groups and the measurement protocol in the Methods section. We also describe the investigated classifiers, the indexes used for performance evaluation, and the experimental design. The findings of the research are presented in the third section. These findings are discussed and criticized in the fourth section. Finally, the fifth section (Conclusion) presents and summarizes the primary outcomes of this research, focusing on the two key proposed objectives.

Results

The biometric and spirometric features of the studied subjects are exhibited in Table 1. The biometric features of the three studied groups were similar, and there were no significant differences between the groups. As can be seen in Table 1, patients with SSc presented significant reductions in the spirometric parameters ($p < 0.05$).

The bar charts in Figure 1 show the characteristics of individuals from the Control Group (CG), Patients with Sclerosis and Normal spirometry (PSNS), and Patients with Sclerosis and Altered spirometry (PSAS). The mean values of each oscillometric parameter were calculated at 95% confidence interval. Using the Analysis of Variance (ANOVA), all oscillometric parameters showed a significant difference in their respective mean values ($p < 0.001$). An increase was observed in the mean values of R_0 , R_m , and Z_r s in patients with Systemic Sclerosis. Thus, disease carriers have higher resistance values (R_0 , R_m) and higher impedance value (Z_r s). On the other hand, resonance frequency (f_r) and the slope of the resistance curve (S) have close values for CG and PSNS group. However, f_r has higher values for PSAS, and more negative values for S . C_{dyn} has higher values for the CG and similar values in the PSNS and PSAS groups.

Figure 2 resumes the results of the experiment 1. One can see that C_{dyn} is the best oscillometric parameter (BOP) to discriminate SSc, presenting moderate diagnostic accuracy ($AUC = 0.77$) for the situation CGvsPSNS and presenting high diagnostic accuracy ($AUC = 0.94$) in the scenario CGvsPSAS. Tables and figures with more detail about these results can be found in the supplement (Tables S1, S2, Figures S1 and S2).

Figure 3 presents the AUCs of the BOP, the ML algorithms, and the MIL classifier obtained in the experiment 2. One can see that the ML algorithms greatly improved the AUC in the situation CGvsPSNS. KNN achieved the best result with $AUC = 0.90$. This result indicates that the algorithm provides a highly accurate diagnose ($0.9 \leq AUC \leq 1.0$). The second-best performance was realized by the ADAB, with $AUC = 0.88$. The comparison of the AUCs with the methodology proposed by DeLong et. al. [38] has shown that KNN, ADAB, RF, and XGB presented a statistically significant difference concerning the BOP.

In the scenario, CGvsPSAS, ADAB, RF, and XGB could provide a small improvement in the AUC. The RF classifier has achieved the best performance with $AUC = 0.97$. RF and XGB exhibited a statistically significant difference regarding the BOP. Table 2 presents the five oscillometric parameters selected in these experiments.

Figure 4 shows the results of the experiment 3, presenting the AUC's obtained by the following strategies: BOP, best ML algorithms with all 7 oscillometric parameters (ML7), best ML algorithm with 5 oscillometric parameters selected by MIL (MIL5+ML), and the best ML algorithm with 5 oscillometric parameters selected by RFE (RFE5 + ML).

Regarding the selection done by MIL, when one compares obtained AUCs with those obtained in the second experiment, it is worth noting that there is only a small decrease in the scenario CGvsPSNS, while in the CGvsPSAS, obtained AUC stays the same. In both situations, the AUCs' comparison with the methodology proposed by Delong et al. has shown a statistically significant difference concerning the BOP. KNN achieves the best AUC in the scenario CGvsPSNS with feature selection done by MIL (AUC=0.87), while in the CGvsPSAS, the best AUC was obtained by RF (AUC=0.97).

The fifth and sixth experiments were designed to train ML algorithms with selection of the three best features. Table 3 presents the selected features, and the Figure 5 resumes the results.

Even with only three features, AUCs' comparison has shown a statistically significant difference concerning the BOP in both cases. In CGvsPSNS, once again, there is a small decrease in the performance (AUC=0.84), and in the CGvsPSAS, the performance was the same (AUC=0.97). Figure 6 presents a 3D picture of the CG and PSNS.

For additional analysis of the ROC curves, Figures 7 and 8 show, respectively, the Se observed at a Sp of 90% and at a Sp of 75% (representing bearable specificity). We included the 90% specificity level inasmuch as it allows only 10% false positives, introducing the most difficult cases into the correct group. It is also noticeable that the sensitivities at 90% Sp of the best ML classifiers were higher than those observed using the BOP in all of the performed experiments. Best ML classifiers invariably presented better results than BOP at 75% Sp.

The interested reader may find a detailed description of the results obtained in the experiments 2 to 6 in the supplement (Tables S3 to S22, Figures S3 to S22).

Discussion

This is the first study on designing an automatic classifier to assist the diagnosis of respiratory abnormalities in patients with SSc using respiratory oscillometry. It was shown that it could simplify the clinical evaluation of lung function and improve these exams' diagnostic accuracy.

In the first experiment, the dynamic compliance (C_{dyn}) was the oscillometric parameter that obtained the best individual performance in both scenarios: CGvsPSNS and CGvsPSAS. In the first scenario, as expected due to the small differences in the measured parameters (Figure 1), it was challenging to separate the Control Group from the Patients with Sclerosis and Normal Spirometry, which yields in AUC=0.77, indicating moderate diagnostic accuracy (Figure 2). In the second scenario, the increase in physiological abnormalities resulted in increased differences in the measured parameters (Figure 1). This allowed C_{dyn} to easily separate the two groups and present an AUC=0.94, which stands for high diagnostic accuracy (Figure 2).

In the second experiment (Figure 3), it is possible to note that the best result for the scenario CGvsPSNS was achieved by the KNN (AUC=0.90). The use of the KNN, ADAB, RF, and XGB algorithms resulted in a

significant improvement in diagnostic accuracy. KNN was followed by ADAB, RF, and XGB using all oscillometric parameters, with ADAB remarkably adjacent to high diagnostic accuracy (AUC=0.88).

Figure 4 consolidates the results for comparison of the AUCs with the selection of five features. In CGvsPSNS, feature selection was not useful for increasing diagnostic accuracy; however, it helped spot the most relevant features. Although the methods selected a different set of features, there is a significant intersection (R0, Rm and Cdyn), which agrees on what can be seen in Figure 1. In the other scenario, CGvsPSAS, RF's best results were achieved, followed by XGB and ADAB (Figure 4). For this scenario, the feature selection has shown that the same results could be achieved using fewer oscillometric parameters. As mentioned in the Introduction, pulmonary manifestation in SSc is characterized by interstitial lung disease associated with pulmonary fibrosis [2]. In this sense, one interesting finding was obtaining Cdyn between the most relevant features in the two studied scenarios (Table 2). This is in close accordance with pathophysiological fundamentals involved in this disease, in which the lungs lose their compliance [2].

Figure 5 summarizes all the results obtained in the fifth and sixth experiments and compares the results in experiments 1 and 2. In CGvsPSNS, the feature selection did not aid in increasing the diagnostic accuracy, but it indicated as important features R0, Zrs and Cdyn, which is in agreement with Figure 1. The use of the reduction of attributes was intended to reduce the complexity of the analysis. The current study found selected features consistent with the presence of lung fibrosis [1, 2]. This rather interesting result is consistent with clinically relevant abnormalities that are known to be associated with reduced survival in these patients [2]. Using the selected three main features, it was possible to inspect the division between groups visually (Figure 6). This optimized interpretation allowed us to observe that the SSc presents smaller values for Cdyn and higher values for R0 and Zrs. Due to its direct physiological translation, this simple spatial description may be beneficial in interpreting the results provided by the proposed medical decision support system, contributing to its use in the clinical scenario.

Concerning the use of the MIL algorithm, it was efficient selecting attributes (Tables 2 and 3), where it was able to achieve a better result for the control versus normal analysis than the one obtained by the specialist selection. However, the MIL algorithm was not as efficient as the traditional classifiers (Figures 3, 4, 5, 7, and 8).

Recent studies have shown the importance of improving our knowledge of the respiratory system [39] and the non-invasive tests of lung function [40-42]. Respiratory oscillometry has been widely perceived as the state-of-the-art in the lung function analysis [43], and one of the most promising emerging technologies in this area [44, 45]. However, although its advantages associated with allowing a simple and detailed examination are particularly important clinically, this method is not yet widely used. One of the main aspects limiting its wide routine application is that the interpretation of the obtained indexes is based on electric models, requiring training and practice. Previous research has established that diagnostic easiness is a fundamental attribute for occupied non-specialist clinicians [46]. The present study contributed to this direction, showing that ML algorithms can improve the medical services

provided to SSc patients, simplifying the use of the respiratory oscillometry and improving the diagnosis of the cited disease.

Early diagnosis of the abnormal respiratory changes in SSc could support early intervention, thus possibly restricting the disease's progression, mitigating adverse symptoms, improving general well-being, restraining complications and comorbidities, and early mortality. Artificial intelligence/machine learning methods have contributed to the improvement of the pulmonary function analysis since the 1980's [5]. The current study extends these findings providing evidence that a combination of respiratory oscillometry and a clinical decision support system based on ML techniques might indicate early abnormal respiratory changes in SSc.

Finally, some important limitations need to be considered and clarified to the reader. First, the study was limited to the Brazilian population at a specific practice site. Thus, it is not possible to ensure its generalizability to a different population. It is suggested that multicenter data be investigated in future studies to expand the generalizability of the findings. It is worth mentioning that by examining the adopted inclusion and exclusion criteria and the present study's biometric features, readers can easily evaluate whether they are likely to achieve similar findings in their own patient population. It is also pertinent to mention that the experimental design of the present work enhances its generalizability. Globally recognized inclusion and exclusion criteria were used, and the work was conducted under usual clinical procedures in a typical setting.

Second, SSc is a disease of low incidence, making it hard to obtain a high number of patients. As a result, the datasets available are relatively small, which requires care to control the complexity of the ML models. In addition to all the care taken in this study to avoid overfitting, such as controlling the hyperparameters, feature selection can also contribute to control overfitting by diminishing the number of inputs. Another reason to employ feature selection is that a smaller number of features can help simplify the analysis. Moreover, if one uses only three features, it is possible to visualize the separation between groups, which can aid the diagnostic explanation.

Conclusions

We designed and tested various classifier methods to achieve a clinical decision support system to assist in detecting respiratory abnormalities in patients with systemic sclerosis. The Respiratory Oscillometry parameters alone can only reach moderate diagnostic accuracy (AUC=0.77) in scenario CGvsPSNS. The ML classifiers' use allowed us to enhance the accuracy, reaching high accuracy (AUC \geq 0.9) in this situation, which represents the initial stages of the disease. In the CGvsPSAS, the oscillometric parameter alone could reach high diagnostic accuracy (AUC=0.94); nevertheless, ML algorithms could provide a small enhancement (AUC=0.97). The developed system may also help simplify the use of oscillometry in the detection of respiratory changes in patients with systemic sclerosis. Notably, the adoption of feature selection has spotted the most crucial oscillometric parameters, which simplify the analysis.

Material And Methods

Eighty-two volunteers were included in the study. Fifty-two, presented SSc and 30 were healthy, composing the control group. The patients with SSc were divided into two groups: (1) the normal spirometry group (n=22), which included patients diagnosed with SSc and showing normal spirometry and (2) the altered spirometry group (n=30) that was composed of patients diagnosed with SSc and presenting altered spirometry, associated with restrictive ventilatory disorder [2].

The exams were conducted at the Pulmonary Function Testing Laboratory of the Pedro Ernesto University Hospital and the Biomedical Instrumentation Laboratory of the State University of Rio de Janeiro. The Hospital Ethical Committee approved the study and all subjects gave informed written consent. This study is in agreement with The Declaration of Helsinki. The criteria for inclusion in the present study were a confirmed diagnosis of SSc according to the consensus of the American College of Rheumatology [9], including volunteers from both genders. The exclusion criteria were a history of exacerbation of disease in the previous 90 days, smoking, and presence of tuberculosis or pneumonia, chronic lung diseases, presence of respiratory infections in the previous 30 days, chest trauma or surgery, inability to perform the tests and chemotherapy and/or radiotherapy for cancer.

The control group was composed by healthy volunteers from both genders without a history of cardiovascular or lung disease or smoking. These individuals do not presented respiratory infections, and showed normal spirometry [10].

The main elements in this study are the respiratory oscillometry measurements, impedance estimation, and clinical decision support system development and performance evaluation. The complete process is shown in Figure 9. Each operation will be described in the next sections.

Respiratory oscillometry measurements and parameters

These analyses were performed using as input excitation small amplitude pressure oscillations ($\leq 2\text{cmH}_2\text{O}$), which were produced by a loudspeaker and applied during tidal breathing at the entrance of the individual's airway through the oral cavity. The result of the exams were generated as the mean of three tests, each 16 seconds long. These tests were considered adequate if they were free of pauses, and presented stable rate and tidal volumes. A pseudo-random noise signal between 4 and 32 Hz was used, and the exams are repeated until all analyzed frequencies presented the minimal coherence function of 0.9. To avoid outlying values, we used a coefficient of variability $\leq 10\%$ in the lowest frequency (4Hz) in the three used tests. The experiments were conducted using an impedance analyzer described previously [11].

A linear regression in the respiratory resistance values in the 4-16 Hz range was used to interpret the obtained results. This yielded resistance at 0 Hz (R_0), the mean resistance in this frequency range (R_m) and the slope of the relationship between the resistive values and frequency (S). The low-frequency range is described by R_0 . This parameter integrates the Newtonian effects, related with the airways, lung and

chest wall resistance, as well as the effect of gas redistribution [12]. The mid-frequency range is described by R_m , which reflects the resistance in the central airways [13]. S is associated with ventilation non-homogeneities [14].

The reactive results were interpreted using four indexes: The mean reactance (X_m), resonance frequency (f_r), the impedance module (Z_{rs}), and the dynamic compliance (C_{dyn}). X_m was calculated using the 4 to 32 Hz frequency range, and describes ventilation inhomogeneity. The f_r occurs when the elastic and inertive properties cancel out, and X_{rs} becomes zero [15]. C_{dyn} was calculated based on the reactance at 4 Hz ($C_{dyn} = 1/2\pi f X_4$), and reflects the respiratory compliance, comprising pulmonary, chest wall, and airway compliances. This parameter is also associated with ventilation homogeneity [13]. Z_{rs} includes the effects of resistance and elastic loads in 4Hz, representing the total mechanical load of the respiratory system [16].

Data sets

In the present work, experiments were executed in a dataset that consisted of 246 measurements acquired from the volunteers. Healthy volunteers contributed with 90 measurements of the oscillometric parameters, patients with sclerosis and normal spirometry with 66, and patients with sclerosis and altered spirometry supplied 90 measurements.

Machine Learning Algorithms

Machine learning algorithms can discover crucial relationships among the features in a set of data [5, 17]. These models' inference can be carried out with minimal user intervention through several techniques such as linear models, graphic models, ensemble strategies, hybrid approaches, and artificial neural networks, among others. In our previous research [6, 7, 18] we have experimented with a wide diversity of models and concluded that ensemble strategies had outstanding performance. In this study, we want to investigate the Extreme Gradient Boosting (XGB) algorithm, a type of ensemble derived from gradient boosting. The final inference model is an assemblage of weak inference models, routinely decision trees. It builds the model in a stepwise mode, where its step is designed to model the error of the previous ones. XGB is an implementation of Gradient Boosting, focusing on regularization to control overfitting, which gives it better performance. Besides, we also want to explore Multiple Instance Learning (MIL) to the early examination of respiratory changes in patients with Systemic Sclerosis. Therefore, in this study, the following ML algorithms were appraised:

- K-Nearest Neighbour (KNN) [19];
- Adaboost with decision trees [20];
- Random Forest (RF) [21];
- Extreme Gradient Boosting(XGB) [22];
- Multiple Instance Learning (MIL) [23];

The first three algorithms have already been briefly described in the previous studies [6, 7, 18]; therefore, we will provide a condensed description of the two algorithms that have not been used in our studies before. A complete description of them can be found in the references.

The Extreme Gradient Boosting is a more efficient, regularized version of Gradient Boosting. In Gradient Boosting, one fits an additive model (ensemble) in a forward manner. In each stage, there is an introduction of a weak learner to cope with the previous weak learners' shortcomings. These shortcomings can be described by the residuals (errors) left by the previous weak learners. Hence, the weak learner to be added must fit the residuals to the ensemble to produce better results. The relation of this algorithm with gradient descent (GD) is since the residuals can be seen as negative gradients, and they can be employed by the GD to locate the minimum value of the loss function. Common choices for the loss function are root mean squared error (regression) and log-loss (classification).

The multi-instance learning (MIL) paradigm was introduced by [23] focused on an application in biochemistry. MIL is considered an extension of supervised learning, where the labels are assigned to a set of instances, known as bags, and not to each instance individually. MIL's central idea is related to the notion of bags: it is labelled as a negative bag (B_i^-) if the total instances contained in it are negative and labelled as positive (B_i^+) if, at best, one of the instances is positive. In this way, a bag can be defined as a collection of instances or regions. The Diverse Density (DD) algorithm was originally introduced by [24], where the algorithm is described as an assessment of the intersection of positive bags minus the union of negative bags. The algorithm's central idea is to find a concept point in the feature space that is close to, at least, one instance of each positive bag and far from the negative bag instances.

Experimental design

This study executed a total of six experiments. The purpose of the first experiment was to investigate the proficiency of a single oscillometric parameter alone to correctly spot the airway obstruction level in patients with systemic sclerosis. We considered two different situations: Control group versus Patients with sclerosis and normal spirometry (CGvsPSNS), and Control group versus Patients with sclerosis and altered spirometry (CGvsPSAS). The remaining experiments also evaluate the two situations described.

The second experiment exploited ML algorithms and compared with the results obtained by a single oscillometric parameter to reveal if the ML algorithms could achieve superior performance. The area under the ROC curve (AUC) was then chosen as the measurement of the performance, since it is regularly employed in medicine [25-28], and yields a superior way to confront classifiers than accuracy [29]. We did not implement feature selection; thus, all of the oscillometric indexes were used. The classifiers described in section 2.3 were realized with Scikit-learn [30], a machine learning library written in python. On the other hand, the Multiple Instance Learning was implemented by the library described in [31]. Since the dataset contains only 246 oscillometric measurements, the k-fold validation procedure [32] is indicated to allow the valuation of the generalization proficiency in the whole dataset. Hyperparameter tuning is a crucial step in model selection. Scikit-learn possesses several strategies to allow hyperparameter fine-tuning,

such as grid search, which experiments with all possible combinations of the hyperparameters. Table J0 presents the classifiers and their respective chosen hyperparameters for tuning.

The third experiment evaluates the capability of MIL as a feature selector with the purpose of complexity reduction and to gain knowledge about the importance of different oscillometric parameters [33]. Its role is to select five oscillometric parameters in a previous step before the classifier training. The fourth experiment employs the recursive feature selection (RFE) also to select five oscillometric parameters before the classifier training. RFE is a wrapper strategy that can use several machine learning algorithms to assess the performance. In this paper, the ML algorithm's choice was the linear support vector machine classifier with L1 regularization. The fifth experiment uses MIL to select three oscillometric parameters, and the sixth employs RFE to choose three oscillometric parameters.

The hypothesis test is a requisite for contrasting ML algorithms. There are a wide variety of parametric tests available, which are commonly based on the t-test [17], [34], [35]. Some of the nonparametric tests most used are McNemar's and Wilcoxon's [34, 36, 37]. In this work, the hypothesis test was carried out with AUCs by applying the methodology specified in Delong et al. [38].

Declarations

Competing interests

The authors declare that they have no competing interests.

Author contributions

DSMA and LMR performed software development, implementation of the computer code and supporting algorithms, analyzed the data, and drafted the manuscript. AJL collected data regarding pulmonary function and provided subject identification and helped to draft the manuscript. JLMA and PLM mentored DSMA, provided funding, participated in the data analysis process and helped to draft the manuscript. All authors read and approved the final manuscript.

Acknowledgements

Not applicable.

Funding

This study was supported by the Brazilian Council for Scientific and Technological Development (CNPq), the Rio de Janeiro State Research Supporting Foundation (FAPERJ) and in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

Availability of data and materials

The datasets used or analyzed during the current study are available from the corresponding author on reasonable request.

Ethics approval and consent to participate

The Research Ethics Committee of the Pedro Ernesto University Hospital granted approval for this study, and the collection and use of the data analyzed in this study.

Consent for publication

All authors consent for the publication of this manuscript.

References

1. Ostojic P, Cerinic MM, Silver R, Highland K, Damjanov N: Interstitial Lung Disease in Systemic Sclerosis. *Lung* 2007, 185(4):211-220.
2. Miranda IA, Dias Faria AC, Lopes AJ, Jansen JM, Lopes de Melo P: On the Respiratory Mechanics Measured by Forced Oscillation Technique in Patients with Systemic Sclerosis. *PLoS ONE* 2013, 8(4):e61657.
3. MacIntyre NR: The Future of Pulmonary Function Testing. *Respiratory Care* 2012, 57(1):154-164.
4. King GG, Bates J, Berger KI, Calverley P, de Melo PL, Dellaca RL, Farre R, Hall GL, Ioan I, Irvin CG, et al: Technical standards for respiratory oscillometry. *The European respiratory journal* 2020, 55(2).
5. Amaral JL, Melo PL: Clinical decision support systems to improve the diagnosis and management of respiratory diseases. In *Artificial Intelligence in Precision Health*. Edited by Barh D. United States: Elsevier; 2020.
6. Amaral JLM, Lopes AJ, Veiga J, Faria ACD, Melo PL: High-accuracy detection of airway obstruction in asthma using machine learning algorithms and forced oscillation measurements. *Computer Methods and Programs in Biomedicine* 2017, 144:113-125.
7. Amaral JLM, Lopes AJ, Faria ACD, Melo PL: Machine learning algorithms and forced oscillation measurements to categorise the airway obstruction severity in chronic obstructive pulmonary disease. *Computer Methods and Programs in Biomedicine* 2015, 118(2):186-197.
8. Amaral JL, Lopes AJ, Jansen JM, Faria AC, Melo PL: An improved method of early diagnosis of smoking-induced respiratory changes using machine learning algorithms. *Computer methods and programs in biomedicine* 2013.
9. Masi AT, Diagnostic SFSCotARA, Committee TC: Preliminary criteria for the classification of systemic sclerosis (scleroderma). *Arthritis & Rheumatism* 1980, 23(5):581-590.
10. Knudson RJ, Lebowitz MD, Holberg CJ, Burrows B: Changes in the normal maximal expiratory flow-volume curve with growth and aging. *American Review of Respiratory Disease* 1983, 127(6):725-734.
11. de Melo PL, Werneck MM, Giannella-Neto A: New impedance spectrometer for scientific and clinical studies of the respiratory system. *Review of Scientific Instruments* 2000, 71(7):2867-2872.

12. Lorino AM, Zerah F, Mariette C, Harf A, Lorino H: Respiratory resistive impedance in obstructive patients: linear regression analysis vs viscoelastic modelling. *The European respiratory journal* 1997, 10(1):150-155.
13. MacLeod D, Birch M: Respiratory input impedance measurement: forced oscillation methods. *Medical & biological engineering & computing* 2001, 39(5):505-516.
14. Brochard L, Pelle G, de Palmas J, Brochard P, Carre A, Lorino H, Harf A: Density and frequency dependence of resistance in early airway obstruction. *Am Rev Respir Dis* 1987, 135(3):579-584.
15. Cavalcanti JV, Lopes AJ, Jansen JM, Melo PL: Detection of changes in respiratory mechanics due to increasing degrees of airway obstruction in asthma by the forced oscillation technique. *Respiratory medicine* 2006, 100(12):2207-2219.
16. Nagels J, Landser FJ, van der Linden L, Clement J, Van de Woestijne KP: Mechanical properties of lungs and chest wall during spontaneous breathing. *Journal of applied physiology: respiratory, environmental and exercise physiology* 1980, 49(3):408-416.
17. Witten IH, Frank E, Hall MA, Pal CJ: *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann; 2016.
18. Amaral JLM, Lopes AJ, Jansen JM, Faria ACD, Melo PL: Machine learning algorithms and forced oscillation measurements applied to the automatic identification of chronic obstructive pulmonary disease. *Computer methods and programs in biomedicine* 2012, 105(3):183-193.
19. Hastie T, Tibshirani R, Friedman J: *The Elements of Statistical Learning*. New York, NY: Springer New York; 2009.
20. Schapire RE: Explaining adaboost. In *Empirical inference*. Springer; 2013: 37-52.
21. Breiman L: Random forests. *Machine learning* 2001, 45(1):5-32.
22. Chen T, Guestrin C: XGBoost: A Scalable Tree Boosting System. In *the 22nd ACM SIGKDD International Conference, 2016*. ACM Press; 2016:785-794.
23. Dietterich TG, Lathrop RH, Lozano-Pérez T: Solving the multiple instance problem with axis-parallel rectangles. *Artificial intelligence* 1997, 89(1-2):31-71.
24. Maron O, Lozano-Pérez T: A Framework for Multiple-Instance Learning. In *Advances in Neural Information Processing Systems 10*. Edited by Jordan MI, Kearns MJ, Solla SA: MIT Press; 1998: 570-576.
25. Hajian-Tilaki K: Receiver Operating Characteristic (ROC) Curve Analysis for Medical Diagnostic Test Evaluation. *Caspian J Intern Med* 2013, 4(2):627-635.
26. Kumar R, Indrayan A: Receiver operating characteristic (ROC) curve for medical researchers. *Indian pediatrics* 2011, 48(4):277-287.
27. Hanley JA, McNeil BJ: The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982, 143(1):29-36.
28. Metz CE: Basic principles of ROC analysis. In; 1978. WB Saunders; 1978:283-298.

29. Jin H, Ling CX: Using AUC and accuracy in evaluating learning algorithms. *IEEE Trans Knowl Data Eng* 2005, 17(3):299-310.
30. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, et al: Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 2011, 12:2825-2830.
31. Tax VDMJC: *MIL, A Matlab Toolbox for Multiple Instance Learning*. 2016.
32. Abu-Mostafa YS, Magdon-Ismael M, Lin H-T: *Learning from data: a short course*. S.l.: AMLbook.com; 2012.
33. Guyon I, Elisseeff A: An introduction to variable and feature selection. *Journal of machine learning research* 2003, 3(Mar):1157-1182.
34. Dietterich TG: Approximate statistical tests for comparing supervised classification learning algorithms. *Neural computation* 1998, 10(7):1895-1923.
35. Alpaydm E: Combined 5×2 cv F test for comparing supervised classification learning algorithms. *Neural computation* 1999, 11(8):1885-1892.
36. Demšar J: Statistical comparisons of classifiers over multiple data sets. *Journal of Machine learning research* 2006, 7(Jan):1-30.
37. Japkowicz N, Shah M: *Evaluating learning algorithms: a classification perspective*. Cambridge University Press; 2011.
38. DeLong ER, DeLong DM, Clarke-Pearson DL: Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988:837-845.
39. Bit A, Chattyopadhay H, Nag D: Study of Airflow in the Trachea of a Bronchopulmonary Patient using CT Data. *Indian Journal of Biomechanics* 2009:31-36.
40. Croxton TL, Weinmann GG, Senior RM, Hoidal JR: Future research directions in chronic obstructive pulmonary disease. *American journal of respiratory and critical care medicine* 2002, 165(6):838-844.
41. Drummond MB, Buist AS, Crapo JD, Wise RA, Rennard SI: Chronic obstructive pulmonary disease: NHLBI Workshop on the Primary Prevention of Chronic Lung Diseases. *Annals of the American Thoracic Society* 2014, 11 Suppl 3:S154-160.
42. Busse WW, Erzurum SC, Blaisdell CJ, Noel P: Executive Summary: NHLBI Workshop on the Primary Prevention of Chronic Lung Diseases. *Annals of the American Thoracic Society* 2014, 11 Suppl 3:S123-124.
43. Bates JHT, Irvin CG, Farré R, Hantos Z: Oscillation Mechanics of the Respiratory System. In *Comprehensive Physiology*. Edited by Terjung R. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2011.
44. Brusasco V, Barisione G, Crimi E: Pulmonary physiology: future directions for lung function testing in COPD. *Respirology* 2015, 20(2):209-218.
45. MacIntyre NR: The future of pulmonary function testing. *Respiratory care* 2012, 57(1):154-161; discussion 161-154.

46. Global Initiative For Chronic Obstructive Lung Disease – UPDATE (2013). “Global Strategy for the Diagnosis, Management, and prevention of Chronic Obstrutive Pulmonary Disease.”

Tables

Table 1 Anthropometric spirometric characteristics of the studied subjects. The last column describes the comparisons between groups, in which the dot means non-significant change, while the dash means significant change.

	Control (1) (n=30)	Normal to the Exam (2) (n=22)	Altered to the Exam (3) (n=30)	
Age (years)	49.7±13.5	9.6±14.5	46.1±12.8	ns
Body mass (kg)	59.2±68.7	60.8±11.9	59.9±13.1	ns
Height (m)	1.6±5.3	1.6±3.4	1.6±6.3	ns
BMI (kg/m ²)	23.5±2.9	24.8±3.9	24.1±4.3	ns
Male/Female	1/29	1/21	1/29	-
FVC (L)	3.4±0.7	2.8±0.7	2.0±0.5	1-2-3-1
FVC (%)	111.9±18.5	96.7±12.5	64.3±11.6	1-2-3-1
FEV ₁ (L)	2.8±0.6	2.4±0.6	1.7±0.5	1-2-3-1
FEV ₁ (%)	112.5±18.2	97.1±11.6	66.4±12.1	1-2-3-1
FEV ₁ /FVC	92.5±10.4	83.4±4.0	86.6±5.5	1-2.3-1
FEF _{25-75%} (L)	3.6±1.0	3.0±0.9	2.5±0.9	1.2-3-1
FEF _{25-75%} (%)	117.4±37.4	112.9±27.2	88.6±28.4	1.2-3-1
FEF/FVC	98.6±28.1	110.0±23.4	134.4±47.8	1.2.3-1

Table 2: Five oscillometric parameters selected by MIL and RFE.

	Control Group versus Patients with Sclerosis and Normal Spirometry (CGvsPSNS)	Control Group versus Patients with Sclerosis and Altered Spirometry (CGvsPSAS)
MIL	Xm, R0, S, Rm, Cdin	fr, R0, Rm, Zrs, Cdyn
RFE	fr, R0, Rm, Zrs, Cdin	Fr,Xm,Rm,Zrs,Cdin

fr: Resonance frequency;

Xm: Mean respiratory reactance;

R0: Respiratory resistance extrapolated at 0 Hz;

S: Slope of the linear relationship of resistance versus frequency;

Rm: Mean respiratory resistance;

Zrs: Absolute value of respiratory impedance in 4 Hz;

Cdyn: Respiratory system dynamic compliance;

Table 3: Three oscillometric parameters selected by MIL and RFE.

	Control Group versus Patients with Sclerosis and Normal Spirometry (CGvsPSNS)	Control Group versus Patients with Sclerosis and Altered Spirometry (CGvsPSAS)
MIL	S, Rm, Cdyn	R0, Zrs, Cdyn
RFE	R0, Zrs, Cdyn	Rm, Zrs, Cdyn

R0: Respiratory resistance extrapolated at 0 Hz;

S: Slope of the linear relationship of resistance versus frequency;

Rm: Mean respiratory resistance;

Zrs: Absolute value of respiratory impedance in 4 Hz;

Cdyn: Respiratory system dynamic compliance.

Figures

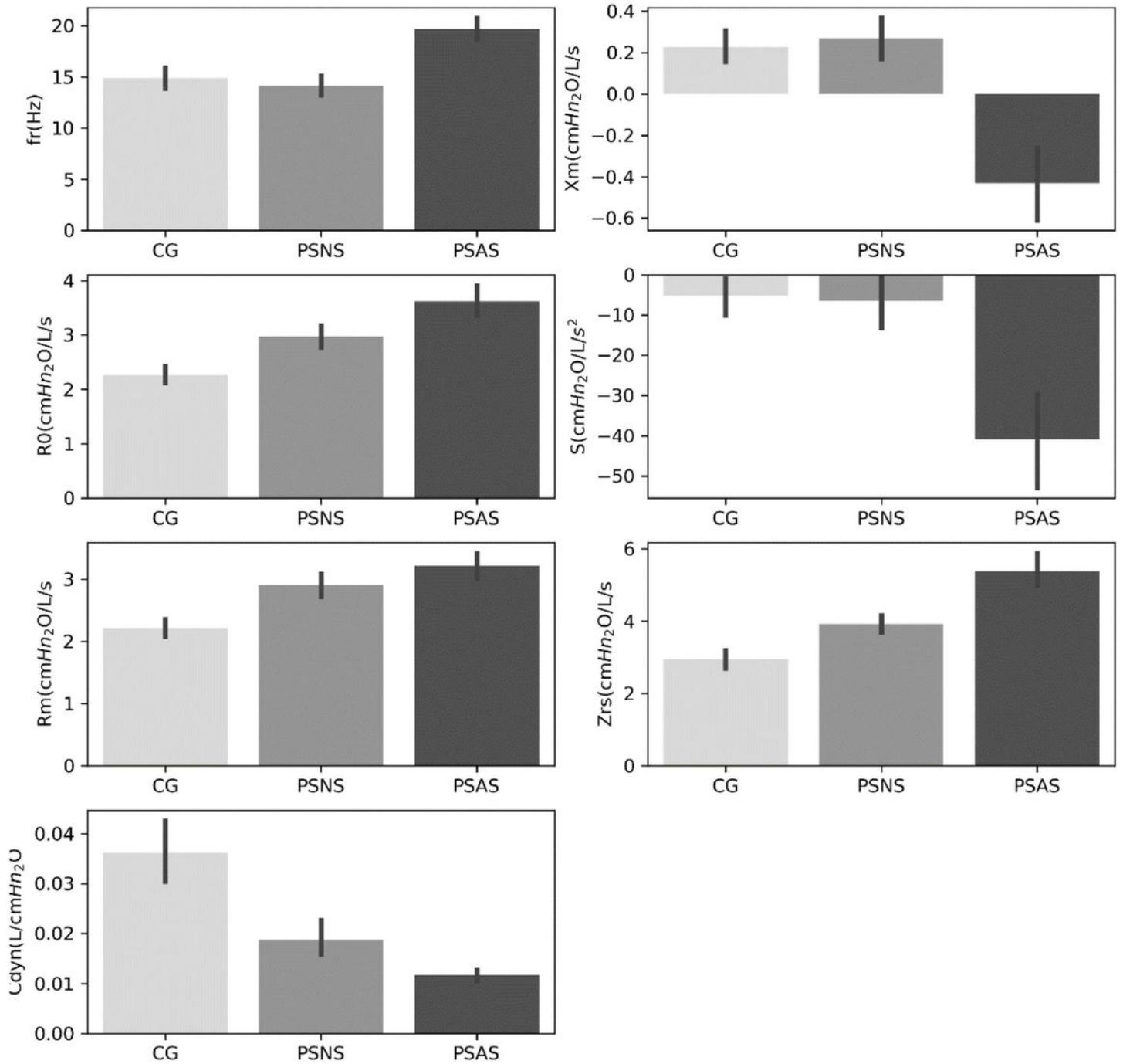


Figure 1

Mean values \pm 95% confidence interval of each FOT parameter. Control Group (CG), Patients with Sclerosis and Normal spirometry (PSNS), and Patients with Sclerosis and Altered spirometry (PSAS). The Analysis of Variance (ANOVA) showed that all parameters presented a significant difference in their respective mean values ($p < 0.001$).

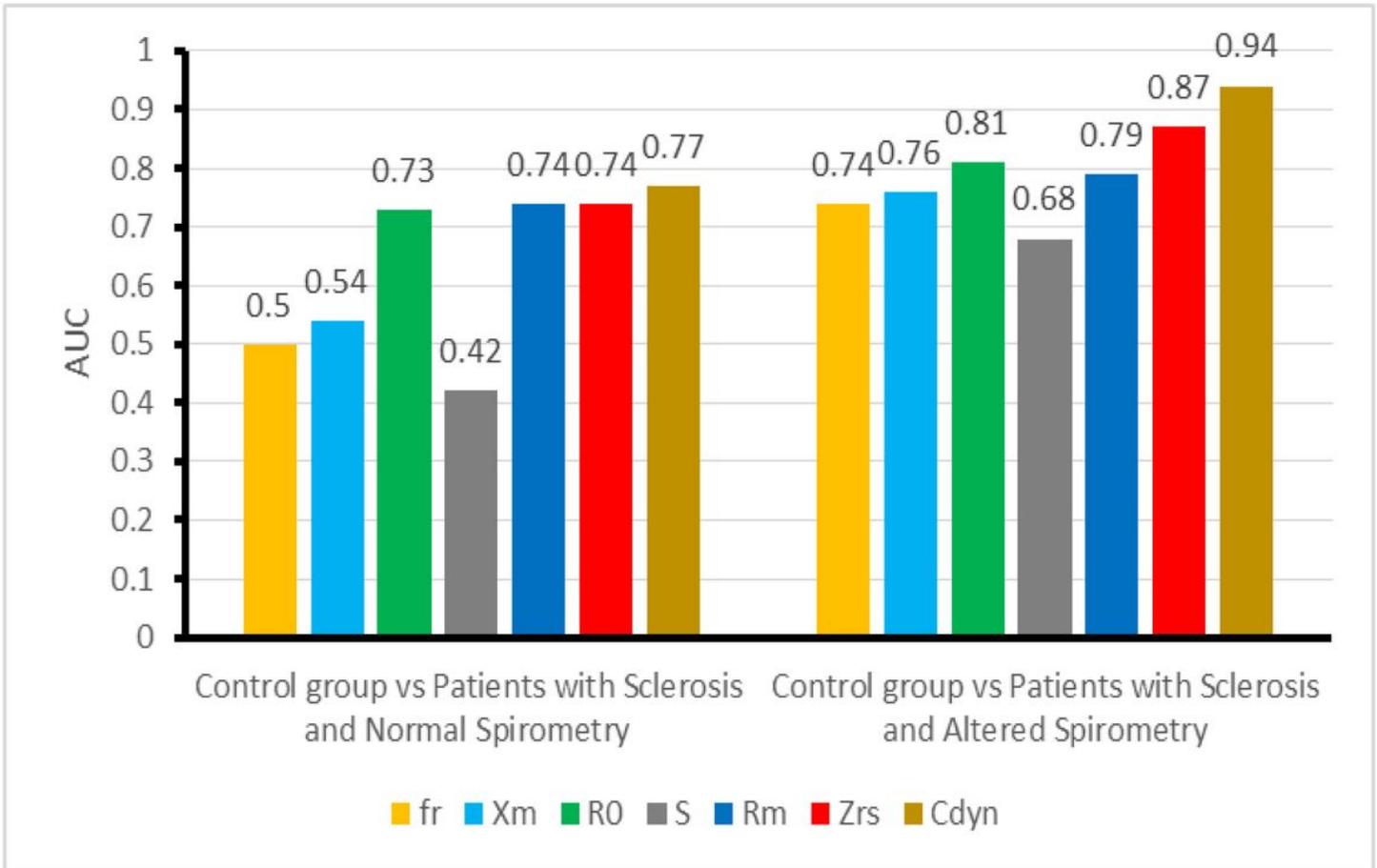


Figure 2

Results of the experiment 1, describing the diagnostic accuracy of oscillometry in sclerosis. fr: Resonance frequency; Xm: Mean respiratory reactance; R0: Respiratory resistance extrapolated at 0 Hz; S: Slope of the linear relationship of resistance versus frequency; Rm: Mean respiratory resistance; Zrs: Absolute value of respiratory impedance in 4 Hz; Cdyn: Respiratory system dynamic compliance.

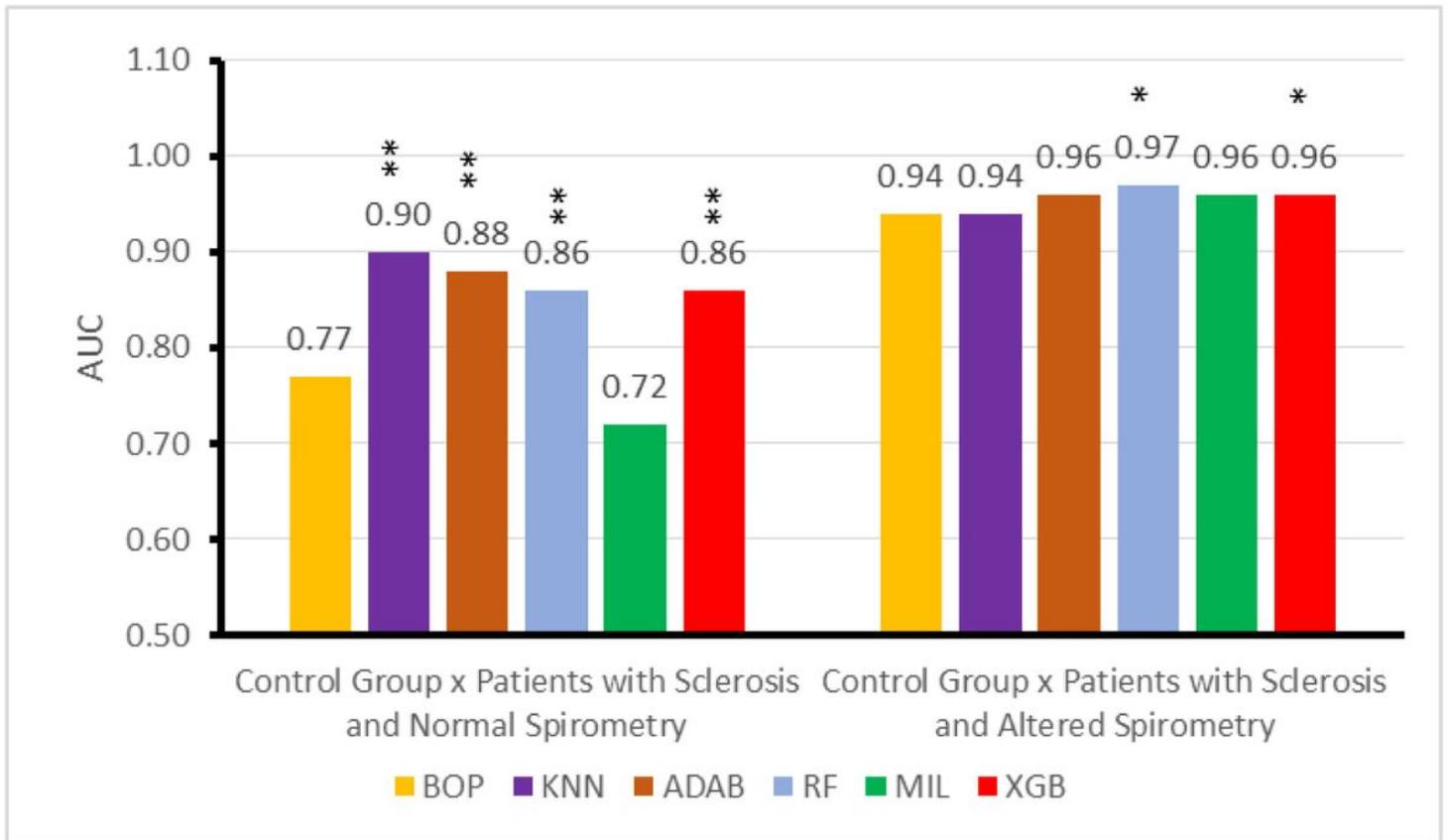


Figure 3

Comparative analysis of the diagnostic accuracy in the experiment 2, considering the best oscillometric parameter (BOP) obtained without use of classifiers), machine learning algorithms and the MIL classifier. K-NN: K Nearest Neighbor; ADAB: Adaboost with decision tree classifiers; RF: Random Forests; MIL: Multiple Instance Learning; XGB: Extreme Boosting Gradient Classifiers; AUC: Area under the ROC curve. Also, “*” indicates that there a statistically significant difference in relation to BOP ($p < 0.05$) and “***” ($p < 0.01$).

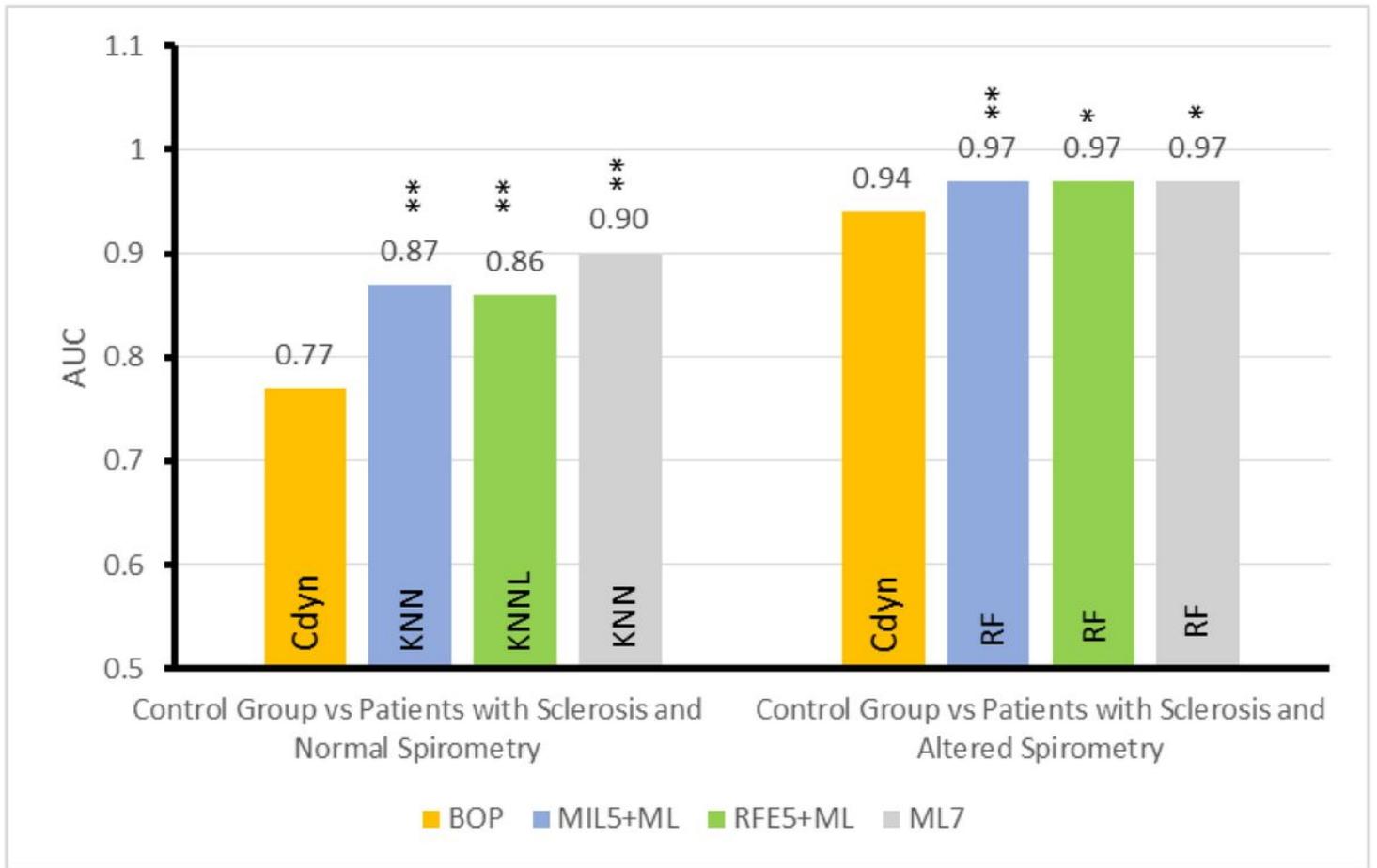


Figure 4

Summary of the Experiment 3 (MIL5+ML: MIL as five feature selector) and Experiment 4 (RFE5+ML: RFE as a five feature selector) – AUCs for the best oscillometric parameter (BOP), for the best ML algorithms in the experiments 3 and 4, and for the best ML algorithm with oscillometric parameters (ML7). The figure indicates the best oscillometric parameter and the best ML algorithm on each case. Also, “*” indicates that there a statistically significant difference in relation to BOP ($p < 0.05$) and “**” ($p < 0.01$).

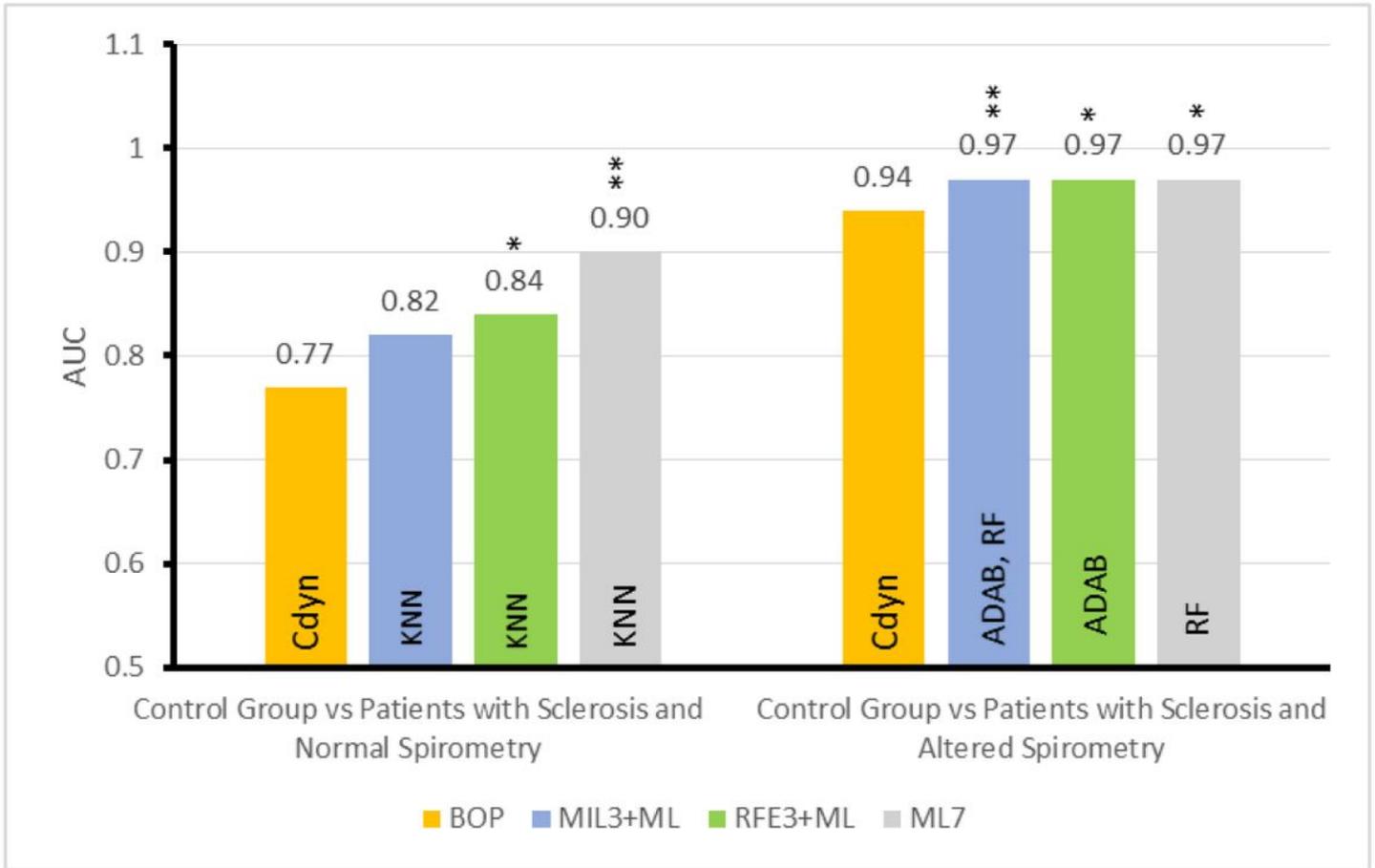


Figure 5

Summary of the Experiment 5 (MIL3+ML: MIL as three feature selector) and Experiment 6 (RFE3+ML: RFE as a three feature selector) – AUCs for the best oscillometric parameter (BOP), for the best ML algorithms in the experiments 5 and 6, and for the best ML algorithm with oscillometric parameters (ML7). The figure indicates the best oscillometric parameter and the best ML algorithm on each case. Also, “*” indicates that there is a statistically significant difference in relation to BOP ($p < 0.05$) and “**” ($p < 0.01$).

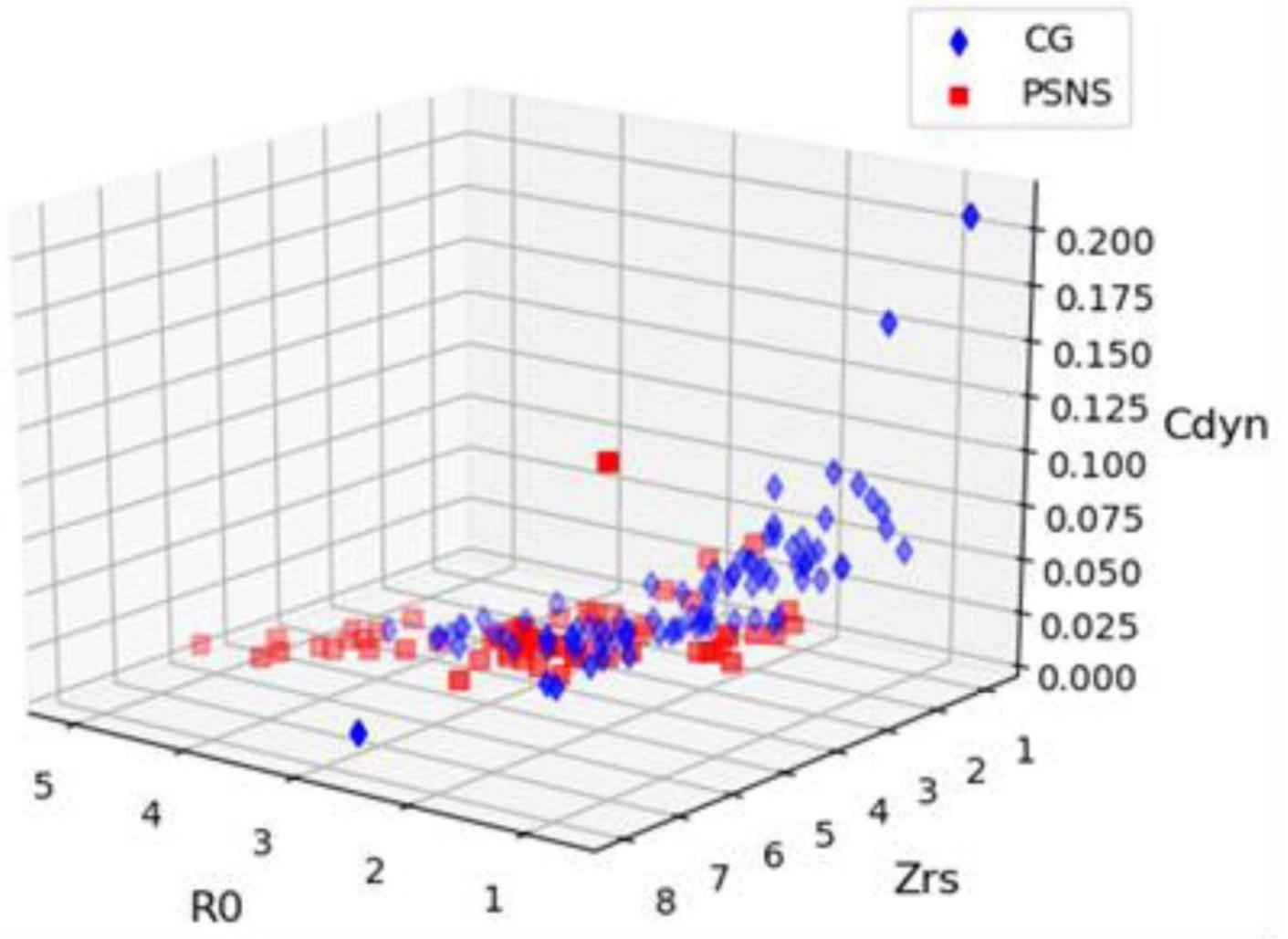


Figure 6

Representation of the Dataset CGvsPSNS using three features: R0, Zrs and Cdyn

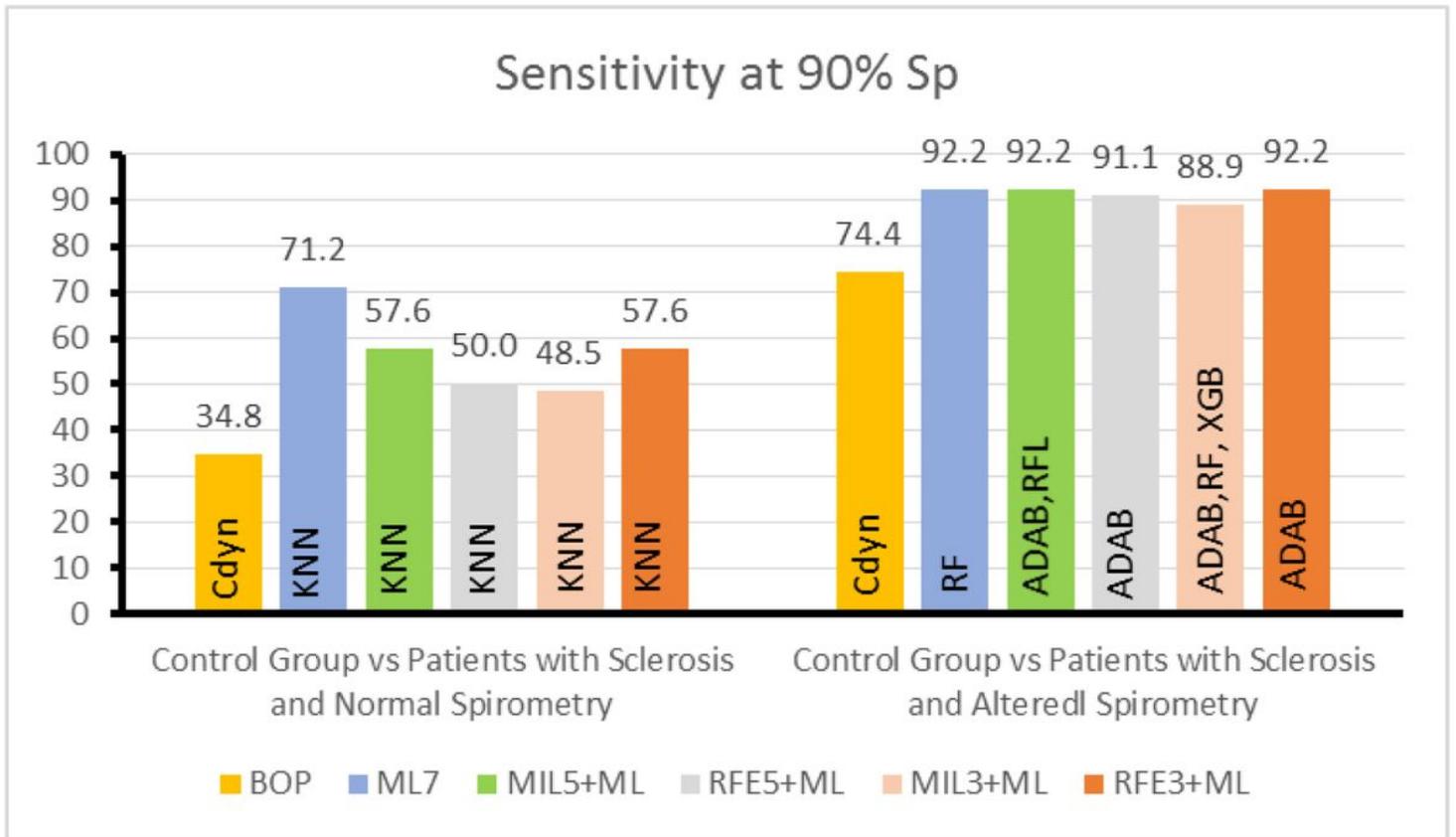


Figure 7

Summary of the experiments describing comparisons of the sensitivity at 90% Sp obtained using the Best Oscillometric Parameter (BOP) and ML methods in all experiments. The sensitivity at 90% Sp presented is that of the best classifier.

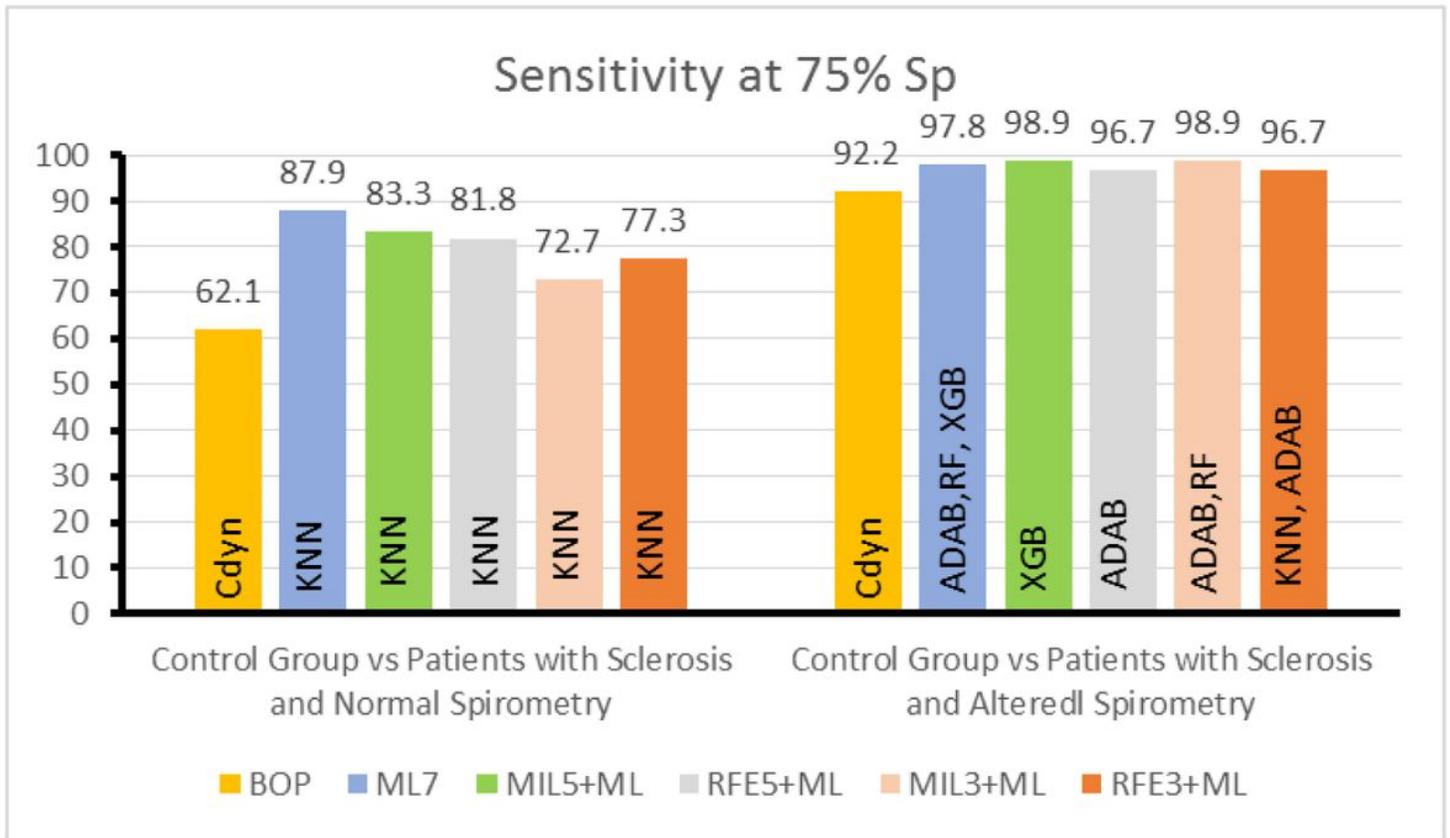


Figure 8

Summary of the experiments describing comparisons of the sensitivity at 75% Sp obtained using the Best Oscillometric Parameter (BOP) and ML methods in all experiments. The sensitivity at 75% Sp presented is that of the best classifier.

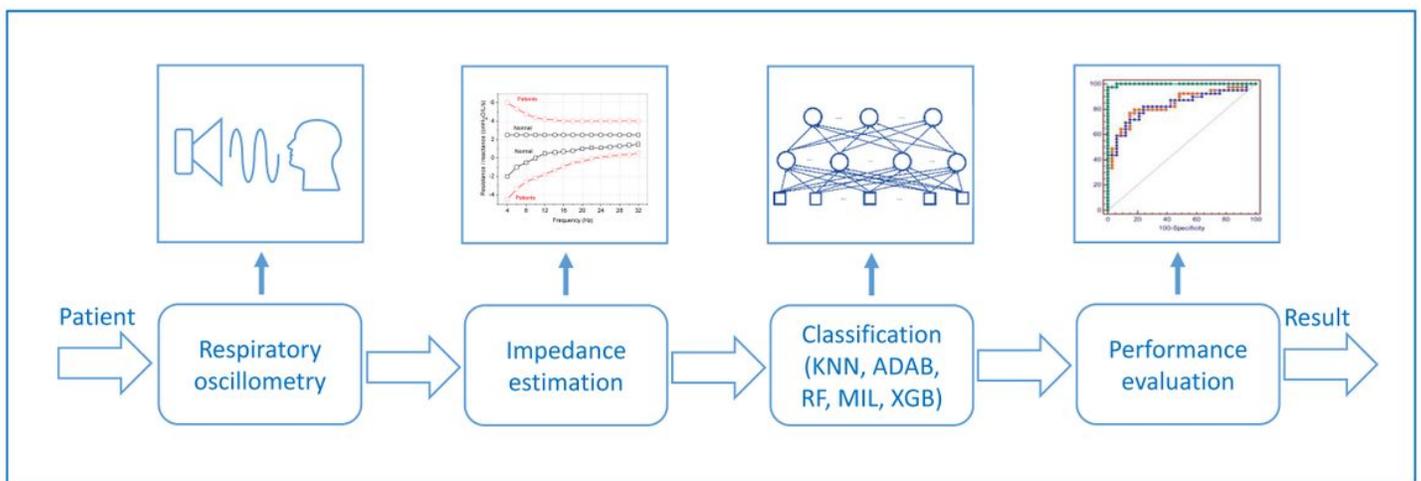


Figure 9

Simplified block diagram describing the main steps in this study. K-NN: K Nearest Neighbor; ADAB: Adaboost with decision tree classifiers; RF: Random Forests; MIL: Multiple Instance Learning; XGB: Extreme Boosting Gradient Classifiers.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [FOTandMLinSystemicSclerosisSupTables06012021.docx](#)