

Article

Assessment of Multi-Layer Perceptron Neural Network for Pulmonary Function Test's Diagnosis Using ATS and ERS Respiratory Standard Parameters

Ahmad A. Almazloum ^{1,*}, Abdel-Razzak Al-Hinnawi ^{2,3}, Roberto De Fazio ⁴ and Paolo Visconti ^{4,*}

- ¹ Department of Electronics and Biomedical Engineering, School of Electrical and Computer Engineering, University of Campinas (UNICAMP), Av. Albert Einstein 400, Cidade Universitária Zeferino Vaz, Campinas 13083-852, Brazil
- ² Faculty of Allied Medical Sciences, Isra University, Amman 11622, Jordan
- ³ Faculty of Science, Isra University, Amman 11622, Jordan
- ⁴ Department of Innovation Engineering, University of Salento, 73100 Lecce, Italy
- * Correspondence: ahmad@ceb.unicamp.br (A.A.A.); paolo.visconti@unisalento.it (P.V.)

Abstract: The aim of the research work is to investigate the operability of the entire 23 pulmonary function parameters, which are stipulated by the American Thoracic Society (ATS) and the European Respiratory Society (ERS), to design a medical decision support system capable of classifying the pulmonary function tests into normal, obstructive, restrictive, or mixed cases. The 23 respiratory parameters specified by the ATS and the ERS guidelines, obtained from the Pulmonary Function Test (PFT) device, were employed as input features to a Multi-Layer Perceptron (MLP) neural network. Thirteen possible MLP Back Propagation (BP) algorithms were assessed. Three different categories of respiratory diseases were evaluated, namely obstructive, restrictive, and mixed conditions. The framework was applied on 201 PFT examinations: 103 normal and 98 abnormal cases. The PFT decision support system's outcomes were compared with both the clinical truth (physician decision) and the PFT built-in diagnostic software. It yielded 92–99% and 87–92% accuracies on the training and the test sets, respectively. An 88–94% area under the receiver operating characteristic curve (ROC) was recorded on the test set. The system exceeded the performance of the PFT machine by 9%. All 23 ATS\ERS standard PFT parameters can be used as inputs to design a PFT decision support system, yielding a favorable performance compared with the literature and the PFT machine's diagnosis program.

Keywords: American Thoracic Society; European Respiratory Society; medical decision support system; multi-layer perceptron neural network; pulmonary function test; respiratory disease



Citation: Almazloum, A.A.; Al-Hinnawi, A.-R.; De Fazio, R.; Visconti, P. Assessment of Multi-Layer Perceptron Neural Network for Pulmonary Function Test's Diagnosis Using ATS and ERS Respiratory Standard Parameters. *Computers* **2022**, *11*, 130. <https://doi.org/10.3390/computers11090130>

Academic Editors: Antonio Celesti, Ivano De Falco, Antonino Galletta and Giovanna Sannino

Received: 5 August 2022

Accepted: 25 August 2022

Published: 29 August 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The Pulmonary Function Test (PFT) is a basic device to evaluate the functionality of the respiratory system [1]. The diagnostic decision obtained from the PFT is usually based on two nonlinear curves: the respiratory flow–volume and volume–time curves. Apart from those primary curves and patient data, many threshold limits and predicted values of the respiratory parameters can be calculated and visualized to the physician to help them make the right decision about the patient condition [2]. The knowledge base from expert pulmonologists could be invested to design a PFT medical decision support system that employs artificial intelligence techniques [3,4], such as an artificial neural network (ANN). An ANN can be trained to deal with sophisticated nonlinear problems [5], such as PFT curves and parameters. Using smart methodologies such as ANN may improve the healthcare quality for patients [3,4].

Many types of ANN have been attempted in the literature for the auto diagnosis of respiratory diseases. This includes Radial Basis Function (RBF) [6–10], MLP [7–17],

Recurrent Networks (RN) [11,13,18], Learning Vector Quantization (LVQ), and Generalized Regression [19]. Different classifiers were utilized, such as Support Vector Regression (SVR) analysis [20], multiclass Support Vector Machine (SVM) [10,21,22], Data Mining [23], Random Forest classifier [22,24,25], and Decision Tree [4,25]. Other attempts combined ANN with different techniques to improve the classifier's performance, such as Fuzzy Algorithms [13,14], Principal Component Analysis [26], k-means clustering [8,22], or using a convolutional neural network to extract the PFT features from the flow time [18]. They all successfully reported a range of performances, indicating 82–100% accuracy.

Of those who utilized ANN with MLP, Baemani et al. [11] implemented eleven PFT parameters as inputs to their two-stage network: the MLP and Elman network, resulting in three-cases classification (normal, obstructive, and restrictive) with an average accuracy of 92.3%. Using 15 inputs to their MLP, Manoharan et al. [7] classified their samples into only two cases (normal or abnormal), achieving 96% classification accuracy. In addition, Jafari et al. [12] obtained 97.6% average classification accuracy by using an MLP neural network with ten inputs, categorizing their samples into four cases (normal, obstructive, restrictive, and mixed). Hakan et al. [17] utilized five PFT parameters to classify their samples into three cases, achieving 98.7% accuracy. Moreover, Badnjevic et al. [13] built a hybrid system using an MLP neural network and fuzzy algorithm to distinguish three cases (normal, COPD (Chronic Obstructive Pulmonary Disease), and asthma) with 99.5% accuracy, utilizing three PFT parameters and two additional measurements. Spathis and Vlamos achieved 89% accuracy differentiating between COPD and asthma by using 13 PFT parameters in their MLP to obtain 89% classification accuracy [22]. The expert system developed by Badnjevic et al. [14] yielded 98.7% correctly classified samples into three classes (normal, COPD, and asthma) with five inputs fed to the ANN. Unlike the others, Loachimescu et al. [15] suggested new input extracted from the area under the expiratory flow–volume curve (AEX). In addition to the conventional three PFT parameters, they were able to classify a relatively large database of samples into four classes (normal, obstructive, restrictive, and mixed) with 91.6% accuracy. Using ten anthropometric features to predict the most important spirometric parameters using MLP, Kalantary et al. [16] were able to correctly establish the patient condition (normal or abnormal), obtaining 84.6% prediction accuracy.

According to the American Thoracic Society (ATS) and the European Respiratory Society (ERS) guidelines [2,27], respiratory diseases can be categorized into three main categories; these are: obstructive disorders (difficulty to exhale), restrictive disorders (difficulty to inhale), and mixed (both obstructive and restrictive disorders) [1]. World Health Organization (WHO) statistics in 2016 revealed that COPD was considered the third world biggest threat after heart diseases and strokes, while lower respiratory infections occupied fourth place [28].

In this study, we assessed the practicability of using the entire respiratory parameters (i.e., 23 features), which are recommended by ATS and ERS in the diagnosis of PFT exams, as inputs to an MLP neural network [29]. In addition, many back-propagation algorithms were used and evaluated and, eventually, compared in this work. To the best of our knowledge, this study has not been investigated. Successfully, our experiment has resulted in up to 99% and 92% accuracy on training and testing sets, respectively. Many standard statistical metrics were utilized for outcome evaluations, including accuracy, sensitivity, specificity, positive predictive value, negative predictive value, and area under the receiver operating characteristic curve.

Additionally, the ANN epoch and the learning rate were recorded. The results were compared with the PFT built-in program, which is often used as a suggestion or a prediction to PFT diagnosis. The suggested PFT medical support system was also compared with similar attempts in the literature.

2. Materials and Methods

Figure 1 illustrates the flow–volume and volume–time curves obtained from the PFT exam. Based on these curves, the PFT machine usually generates a set of respiratory parameters called spirometric measurements/parameters [30]. The major parameters are: Forced Expiratory Volume (FEV₁), Forced Vital Capacity (FVC), ratio FEV₁/FVC (briefly FEV₁%), Forced Expiratory Flow (FEF), Peak Expiratory Flow (PEF), and Maximum Expiratory Flow (MEF). Other parameters are acquired at different time intervals, such as FEV₁, which indicate the forced expiratory volume during the first second, or at different percentages such as MEF₂₅%, which indicates 25% maximum expiratory flow at 75% of FVC [2,27]. Table 1 shows the full set of the 23 respiratory parameters specified by ATS and ERS [2,27]. Usually, those parameters are classified as either the essential respiratory ones (Group 1) or the secondary respiratory parameters (Group 3). Instead, Group 2 in Table 1 illustrates the patient anthropometric characteristics, including age, weight, height, sex, and smoking/nonsmoking. In total, a set of 23 PFT features are usually recorded in the PFT machine according to ERS/ATS recommendations. The pulmonologists observe these PFT parameters in comparison to deduce diagnosis.

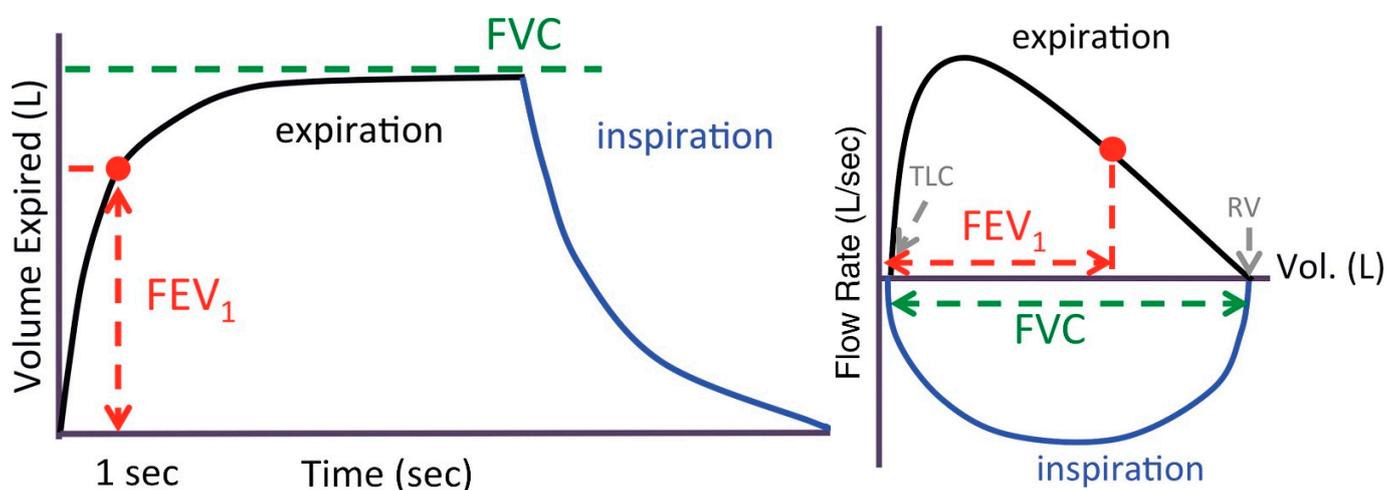


Figure 1. Graphical representation of the volume flow (left) and flow time (right), the spirometric measurements [29]. FVC—Forced Vital Capacity, FEV₁—Forced Expiratory Volume at the 1st second, TLC—Total Lung Capacity, RV—Residual Volume.

Table 1. List of the ATS/ERS standard respiratory parameters in the PFT exam. All the abbreviations are explained in Section 2.

Group	Parameters
Group 1	FVC, FEV ₁ , (FEV ₁ /FVC) %, FEF (25–75%)
Group 2	Age, Weight, Height, Sex, Smoking/non-smoking
Group 3	PEF, MEF ₂₅ %, MEF ₅₀ %, MEF ₇₅ %, FEV _{0.5} , FEV ₂ , FEV ₃ , PEFT, FEV ₂ /FVC, FEV ₃ /FVC, FEF _{25–50} %, FEF _{50–75} %, FEF _{75–85} %, FEF _{0.2–1.2}

The PFT device is usually supplied with built-in diagnostic software, suggesting suspicion of the presence of respiratory disorders. In this experiment, the PFT software was designed by COSMED Cardiopulmonary Diagnostics, Italy [31]. It suggests the diagnosis of the patient’s status, namely “normal” or “abnormal” with the suspicion of obstructive, restrictive, or mixed conditions. The software employs an IF/THEN program based on the comparison between the recorded and typical values. The software utilizes the ATS/ERS standard respiratory parameters (Table 1).

We collected 201 retrospective PFT examinations from a pulmonary function device (Quark PFT device, manufactured by COSMED Cardiopulmonary Diagnostics, Pavona

RM, Italy [31]) in a university hospital archive relating to a recent three-year period (just before the SARS-CoV-2 pandemic). Personal information was removed, so they became anonymous medical records. For every PFT exam, we had the pulmonologist's diagnostic decision of whether the patient condition was normal or suspicion of restrictive, obstructive, or mixed disorders. There were 103 normal cases (51%), 74 obstructive (37%), 14 restrictive (7%), and 10 mixed (5%) cases. Based on the 75–25% strategy, the 201 samples were randomly divided into training and test sets. It is worth mentioning that a bigger dataset of 381 samples with 13 features was also tested and evaluated (results not shown), but it did not outperform the 201 dataset with 23 features. Thus, the latter (with 201 samples) was only considered in this work. Additionally, the 201 dataset was filtered carefully from a larger dataset (consisting of about 1000 samples), and only samples that followed the ATS/ERS standards and were certified by the pulmonologists were included. Table 2 illustrates the distribution of the 201 samples in the experimental dataset. All experiments were tested and evaluated in a MATLAB environment program version R2017b (Math Works. Inc., Natick, MA, USA), utilizing its artificial neural network library [32].

Table 2. The distribution of the dataset.

	Normal	Obstructive	Restrictive	Mixed
Training (75%)	76	55	8	7
Test (25%)	26	19	6	3
Total	102	74	14	10

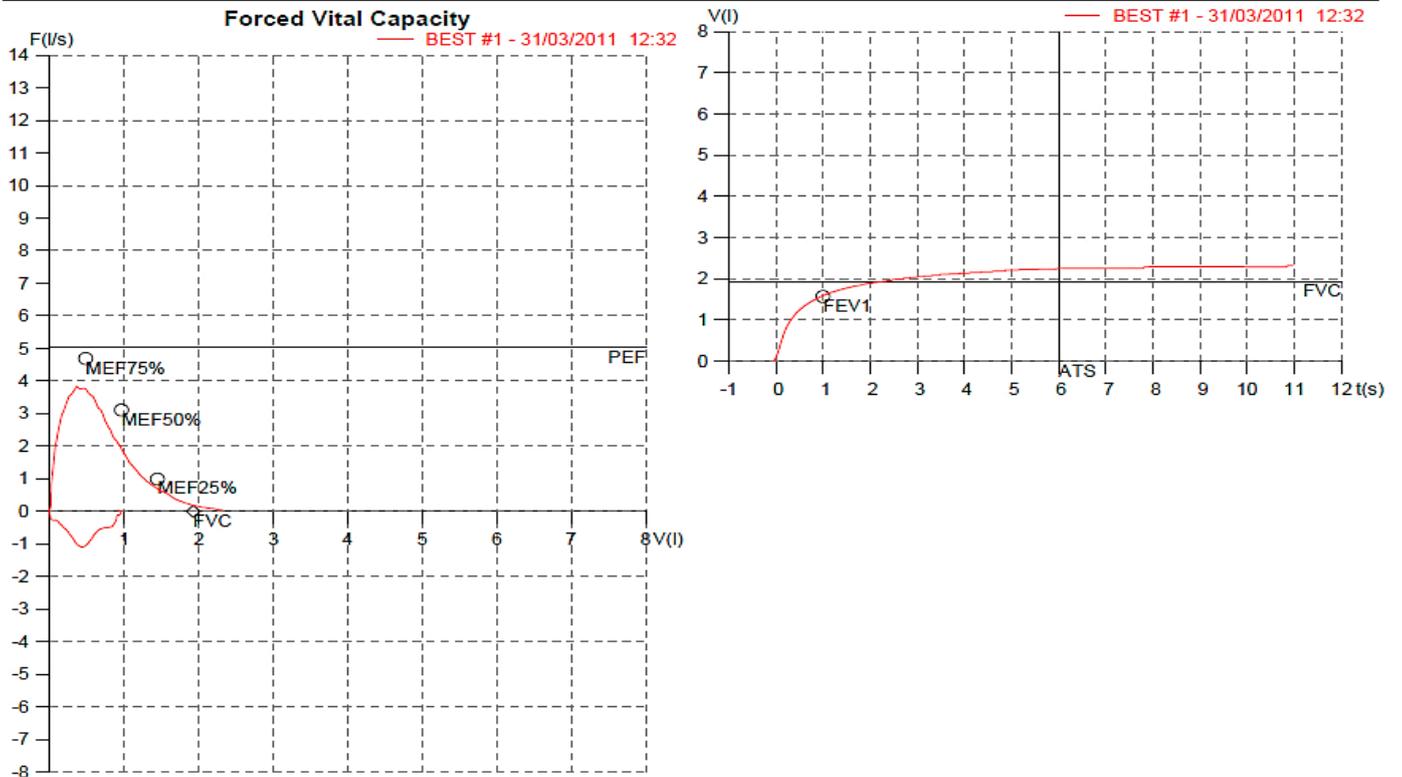
Even though the used dataset is relatively small compared with other works in the literature, it is certified and confinable. Instead, other works in the literature had a bigger dataset, but this was not enough to improve the accuracy. In other words, not only the quantity of the samples matters but also the quality (i.e., a variability that means distinct samples covering a wide range of the input space).

Figure 2 illustrates a sample from the 201 Pulmonary Function Test (PFT) exams that have been analyzed in this research work. It shows the anthropometric characteristics of the patient (at the top of Figure 2), the flow–volume curve as “Forced Vital Capacity (FVC)” (on the left center), the volume–time curve as “Forced Expiratory Volume (FEV)” (on the right center), and the summary report (at the bottom of Figure 2) as a table for the 13 respiratory parameters that are usually calculated from these two curves. In this summary report, the predicted values (Pred.) of each PFT parameter are estimated according to the patient anthropometric characteristics (4th column). Instead, the actual measurement of the PFT parameter (“BEST#1”) and its percentage (“%Pred.”) related to the predicted value are shown in the 5th and 6th columns, respectively.

The final diagnosis suggested by the built-in PFT program is reported at the bottom of the report as “Normal Spirometry”. The MEF75%, MEF50%, MEF25%, FVC, PEF, and FEV1 values, which are reported on the curves, are examples of the respiratory parameters. It is worth mentioning that the patient maneuver, as shown in the volume–time curve, should last for at least 6 s according to the American Thoracic Society (ATS) and the European Respiratory Society (ERS) to obtain reliable results.

Table 3 illustrates the descriptive statistics of all the 201 Pulmonary Function Test (PFT) samples for the 23 PFT parameters. It shows the mean and the standard deviation (SD) of all 23 respiratory parameters, which are recommended by the American Thoracic Society (ATS) and the European Respiratory Society (ERS), for the 201 PFT exams. For the Sex parameter, the number of men and the related percentage were calculated instead of the mean and SD values. Likewise, the number of positive cases of smoking and the related percentage in the database were calculated for the Smoker parameter.

Last Name: XXXXXXXXX **Date of Birth:** 05/08/1943 **Age:** 67
First Name: XXXXXXXXX **Sex:** XXXXXXXXX **Weight (Kg):** 59.0
ID: XXXXXXXXX **Ethnic Corr.:** XXXXXXXXX **Height (cm):** 148.0
Date: 31/03/2011 **Description:** **BMI (Kg/m²):** 26.9
Predicted: ERS 93 **Company:** **Smoke:** No



Parameter	UM	Description	Pred.	BEST#1	%Pred.
Best FVC	l(btps)	Best Forced Vital Capacity	1.92	2.34	121.6
FVC	l(btps)	Forced Vital Capacity	1.92	2.34	121.6
FEV1	l(btps)	Forced Exp Volume in 1 sec	1.57	1.60	101.8
PEF	l/sec	Peak Expiratory Flow	5.02	3.84	76.5
FEV6	l(btps)	Forced Exp Volume in 6 sec	2.32	2.28	98.3
FEV1/FVC%	%	FEV1 as % of FVC	76.4	68.3	89.5
FEV6/FVC%	%	FEV6 as % of FVC		97.3	
FEV1/FEV6%	%	FEV1 as % of FEV6		70.2	
FEF25-75%	l/sec	Forced mid-expiratory flow	2.49	0.98	39.2
MEF75%	l/sec	Max Exp Flow @ 25% FVC	4.69	3.56	75.9
MEF50%	l/sec	Max Exp Flow @ 50% FVC	3.11	1.32	42.5
MEF25%	l/sec	Max Exp Flow @ 75% FVC	0.99	0.34	34.3
FET100%	sec	Forced Expiratory Time		10.9	
PEFr	l/min	Peak Expiratory Flow (l/min)	301.2	230.3	76.5

Diagnosis:
 Normal Spirometry
 Printed 01/03/2012
 PFT Suite 9.1b

Figure 2. An Example of the PFT curves (i.e., the print-out report) from the COSMED-Quark device. A sample from the dataset employed in this research work.

Table 3. The mean and SD values of the 201 samples of all PFT parameters.

#	PFT Examination Parameters	Normal 103 Samples		Obstructive 74 Samples		Restrictive 14 Samples		Mixed 10 Samples	
		Mean \pm	SD	Mean \pm	SD	Mean \pm	SD	Mean \pm	SD
1	FVC (L)	4.23	0.91	3.31	0.70	2.68	0.45	2.67	0.39
2	FEV1 (L)	3.14	0.72	1.91	0.49	2.14	0.40	1.93	0.37
3	(FEV1/FVC)%	74.06	5.36	57.45	5.68	79.71	6.10	71.56	7.96
4	FEF25–75% (L/min)	2.58	0.96	0.98	0.33	2.52	1.15	1.57	0.60
5	Age (year)	50.08	11.51	55.85	10.35	46.93	12.22	51.60	12.60
6	Height (cm)	167.72	7.74	166.14	5.93	164.86	8.63	170.70	7.10
7	Weight (kg)	78.70	12.73	76.43	14.59	73.36	16.41	75.30	10.58
8	Sex (men/%)	80.00	77.7%	64.00	86.5%	11.00	78.6%	8.00	80.0%
9	Smoking (+/%)	7.00	6.8%	8.00	10.8%	3.00	21.4%	1.00	10.0%
10	PEF (L/min)	7.10	1.65	4.67	1.16	6.43	1.71	5.08	0.82
11	MEF25% (L/min)	1.10	0.52	0.41	0.14	1.08	0.54	0.72	0.37
12	MEF50% (L/min)	3.32	1.11	1.23	0.43	3.16	1.39	1.91	0.75
13	MEF75% (L/min)	5.90	1.30	2.63	0.87	5.56	2.09	3.76	1.32
14	FEV0.5 (L)	2.35	0.53	1.34	0.35	1.74	0.37	1.47	0.34
15	FEV2 (L)	3.67	0.83	2.47	0.60	2.41	0.43	2.29	0.38
16	FEV3 (L)	3.89	0.86	2.77	0.64	2.52	0.46	2.44	0.37
17	PEFT (s)	98.69	32.50	66.50	44.65	78.95	32.33	67.94	30.88
18	(FEV2/FVC)%	86.59	4.16	74.46	4.80	89.74	3.56	84.86	5.85
19	(FEV3/FVC)%	91.86	3.20	83.54	4.07	93.75	2.66	90.84	4.28
20	FEF25–50% (L/min)	4.42	1.22	1.74	0.59	4.24	1.75	2.64	1.04
21	FEF50–75% (L/min)	1.86	0.78	0.69	0.24	1.82	0.88	1.15	0.49
22	FEF75–85% (L/min)	0.72	0.37	0.31	0.10	0.63	0.33	0.49	0.26
23	FEF0.2–1.2 (L/min)	6.07	1.57	2.78	1.24	5.34	1.92	3.42	1.46

2.1. Spirometer Procedure

All patients were instructed to perform the standard spirometry procedure recommended by the ATS/ERS [2]. In detail, the patient inhaled rapidly and completely, then exhaled in the mouthpiece (turbine flow meter) forcefully until no more air could be expelled, keeping the nose clip attached and the mouth tightly closed on the mouthpiece. The procedure was repeated until acceptable results were obtained, and the best outcomes were stored for further analysis.

2.2. MLP Neural Network

The experiment employed the Multi-Layer Perceptron Neural Network (MLPNN or MLP for short) with a back-propagation (BP) algorithm [5,32], as shown in Figure 3, also known as a fully connected feed-forward back-propagation neural network. This MLP is basically made up of an input layer, an output layer, and at least one hidden layer. Each layer consists of a high number of neurons, based on the layer location and ruled by essential constraints. For example, the number of neurons in the input layer should be determined by the number of sample features, which are the 23 features in this experiment. The number of neurons in the output layer is ruled by the number of decisions that the ANN will classify (e.g., normal, obstructive, restrictive, and mixed). Whereas, for hidden layers, the at least maximum number of neurons can be calculated according to Equation (1) explained by Stathakis et al. [33–35], where “ N ” is the number of distinct samples and “ m ” is the number of output neurons. Selecting higher limits than what theoretically can be calculated from Equation (1) would guarantee the optimum combination of neurons during training [35]. The number of neurons in the hidden layers was varied during the

training tryouts (searching for the optimum performance) from 2 to 65 in the first and from 2 to 32 in the second for each individual back-propagation algorithm [33]. Thus, $32 \times 65 = 2080$ structures of NNs were constructed/trained and then evaluated for each algorithm. The optimum number of neurons was determined after several variations until the ANN yielded a minimum error (i.e., highest performance) [5,32].

$$L_1 = \sqrt{(m+2)N} + 2\sqrt{N/(m+2)}, \quad L_2 = m\sqrt{N/(m+2)} \quad (1)$$

Input Features [-1, +1]

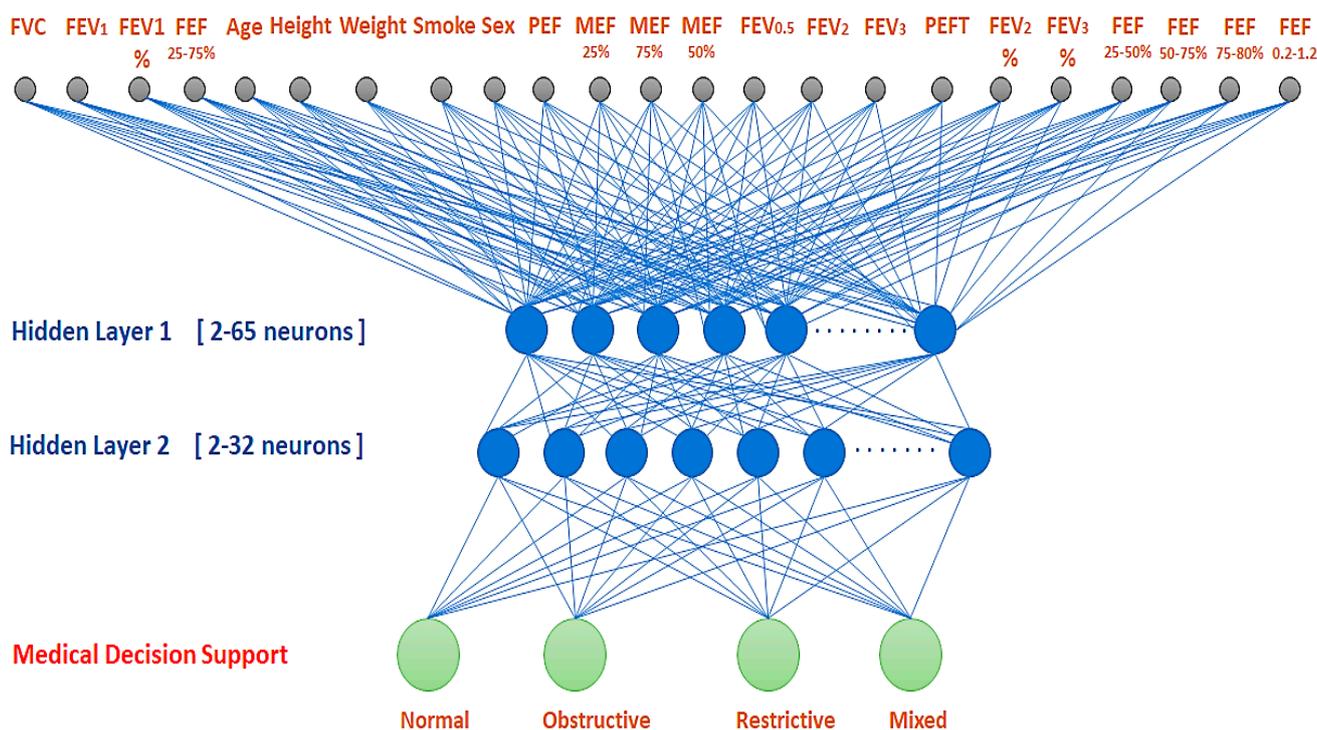


Figure 3. The MLP medical decision support architecture for PFT exam using the 23 respiratory parameters from ERS/ATS. The abbreviations are explained in detail in Section 2.

The MLP architecture was initially tested several times by changing the normalization scale of the input data (e.g., $[-1, +1]$ and $[0, +1]$), the type of activation function, and the number of hidden layers (e.g., 1 or 2 layers). After these preliminary tryouts, the best performance was achieved by employing a $[-1, +1]$ normalization scale and two hidden layers. A single hidden-layer neural network may be sufficient in any decision support system, but a complex activation function should be used [36]. The second hidden layer is important to compensate for any loss in network performance [34–36]. Thus, we utilized two hidden layers. Increasing the hidden layers would require an increment in time and processing capacity. Then, the nonlinear symmetric sigmoidal activation functions were implemented in all neurons of the input, hidden, and output layers. This function was preferred because they achieved the best performance during the preliminary tryouts. The early stopping criterion was employed; namely, the training immediately stops once the validation error starts to increase [32]. Finally, the proposed MLP architecture (Figure 3) was trained and then tested for all the 13 possible BP algorithms to validate the research work objective, namely the investigation of the validity of using all 23 ARS and ETS PFT parameters to design a PFT medical decision support system. The planned MLP PFT medical decision support architecture is illustrated in Figure 3.

2.3. Measures of Classification Performance

There are well-established statistical metrics for assessing the classification performance of medical decision support systems [37]. In our research work, we employed six different statistics parameters [37,38]; these are: the classification accuracy, sensitivity, and specificity. Additionally, the positive predicted value (PPV) and negative predicted value (NPV) are other statistical indicators. Equations (2)–(7) explain these metrics, considering that TP , TN , FP , and FN represent the true positive, true negative, false positive, and false negative cases, respectively. The variations in sensitivity versus specificity generate the receiver operating characteristics (ROC) curve, which is another statistical indicator for the ANN classification performance resulting in another indicator, the area under the curve (AUC) (Equation (7)). Further ANN model statistical indicators could have been utilized, such as those reported in Ref. [16]. However, Equations (2)–(7) are commonly used in assessing the classification performance of the ANN bioinformatics applications [37–39]. It is important to note that many performance metrics (namely sensitivity, specificity, positive predictive value, negative predictive value, ROC, and area under ROC), besides the accuracy, were used to evaluate the classification outcomes well in a reliable and consistent manner, taking into consideration the imbalanced classes in the ANN output [40]. Using different measures of accuracy helps to reveal blind spots, to avoid misleading evaluation, and to provide many accurate outcomes [41]. In addition, as the other metrics have provided high percentages, the classifier's performance was considered reliable.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

$$Sensitivity = \frac{TP}{TP + FN} \quad (3)$$

$$Specificity = \frac{TN}{TN + FP} \quad (4)$$

$$PPV = \frac{TP}{TP + FP} \quad (5)$$

$$NPV = \frac{TN}{TN + FN} \quad (6)$$

$$AUC = \int_0^1 ROC(t) dt \quad (7)$$

3. Results

Several experiments that we have already done, using fewer parameters (e.g., 13 features/inputs) with the same strategy explained in Section 2.2, resulted in worse accuracies than using the same number of samples with 23 features/inputs. Thus, we concluded that the dataset with more features outperformed the one with fewer features, remembering that all 23 features were recommended by the international standards (ATS/ERS) for diagnosis. After fixing the number of features, we started many further experiments to determine other hyperparameters related to the neural network (NN), which is the central focus of this research work (several training algorithms).

In the first experiment, the 23 spirometer measurements from all training samples in Table 2 were fed to the proposed MLP architecture shown in Figure 3. The mean and SD of the values of all parameters are shown in Table 3. After training, the MLP was evaluated on samples in the test set. The classification accuracy was reported on both the training and test sets for each BP algorithm. Table 4 reports the obtained results and the optimum number of neurons in the hidden layers, the epoch, and the learning rate.

Table 4 demonstrates that the fixed number of neurons in each hidden layer corresponds with the best results obtained after training and parameter tuning. Thus, no encoder/decoder structure of the recurrent neural network was needed; instead, we used

a fully connected feed-forward back-propagation neural network or Multilayer Perceptron Neural Network (MLP).

Table 4. The MLP architecture and classification accuracy on the training and test sets for different 13 BP algorithms. LR—learning rate, NA—not applicable.

#	BP Algorithm	No. of Neurons Hidden Layer 1	No. of Neurons Hidden Layer 2	Epoch	LR	Accuracy (Training)	Accuracy (Test)
1	Levenberg Marquardt (LM)	7	21	21	<0.01	0.99	0.90
2	Bayesian Regularization (BR)	6	24	20	2.36	0.96	0.89
3	Resilient Back Propagation (RBP)	47	16	60	0.01	0.98	0.89
4	Scaled Conjugate Gradient (CGS)	47	28	55	<0.01	0.96	0.87
5	Polak–Ribiere Conjugate Gradient (CGP)	36	18	47	0.01	0.97	0.87
6	Powell–Beale Conjugate Gradient (CGB)	45	30	42	0.01	0.97	0.92
7	Fletcher–Powell Conjugate Gradient (CGF)	43	19	31	0.01	0.96	0.90
8	One Step Secant (OSS)	4	28	43	0.01	0.96	0.89
9	Gradient Descent with Momentum and Adaptive Learning Rate Rule (GDX)	50	15	158	2.62	0.96	0.88
10	Gradient Descent with Adaptive Learning Rule (GDA)	64	4	212	0.68	0.94	0.90
11	Gradient Descent (GD-1)	24	29	1000	0.01	0.92	0.87
12	Sequential Order Incremental Training with Learning Functions (SOIT)	25	10	1000	NA	0.92	0.88
13	Batch Training with Weight and Bias Learning Rules (BT)	31	30	1000	NA	0.92	0.87

To carry on the experiments, we preferred six BP algorithms from the 13 BP detailed in Table 4. The CGB was selected because it led to the highest accuracy (i.e., 97% and 92% accuracy on the training and test sets, respectively) among the four possible Conjugate Gradient (CG) algorithms, the CGS, CGP, CGB, and CGF. The LM algorithm was often reported in the literature [7–15], so it was particularly considered; it yielded the highest accuracy (i.e., 99%) on the training set. The BR, RBP, and the OSS were selected because they performed high accuracy on both the training and the test sets. Finally, the GDA was preferred because it yielded to the best accuracy among the three Gradient Descent (GD) algorithms, the GDX, GDA, and GD-1. All these BP algorithms are highlighted in Table 4. Subsequently, in the second experiment, we evaluated the MLP performance for the selected six BP algorithms in terms of sensitivity, specificity, PPV, NPV, and AUC. This objective was achieved on the test set samples by comparing the MLP decision with the clinical decision (i.e., the pulmonologist’s diagnostic decision about patient condition). Table 5 illustrates the obtained results.

Finally, the third experiment reported the MLP’s accuracy of the selected six BP algorithms in terms of each MLP’s output (i.e., the decision whether the PFT was normal, obstructive, restrictive, or mixed). This experiment was attained on all samples and the

training and test sets (i.e., the 201 samples in Table 2). Consequently, the MLP's accuracy was possibly compared with the PFT IF/THEN program's accuracy, which was in the same manner estimated from the 201 samples. Figure 4 demonstrates the results; the PFT device outputs are represented by green bars while the MLP outputs are represented by different colored bars.

Table 5. Statistical results of MLP neural network in comparison with the clinical truth. PPV—positive predictive value, NPV—negative predicted value, AUC—area under the receiver operating characteristic curve (ROC).

#	BP Algorithm	Accuracy	Sensitivity	Specificity	PPV	NPV	AUC
1	Levenberg Marquardt (LM)	0.90	0.71	0.92	0.73	0.92	0.88
2	Bayesian Regularization (BR)	0.89	0.76	0.92	0.55	0.91	0.94
3	Resilient Back Propagation (RBP)	0.89	0.54	0.92	0.50	0.91	0.93
4	Powell–Beale Conjugate Gradient (CGB)	0.92	0.81	0.94	0.63	0.93	0.93
5	One Step Secant (OSS)	0.89	0.57	0.92	0.55	0.91	0.94
6	Gradient Descent with Adaptive Learning Rule (GDA)	0.90	0.77	0.93	0.57	0.92	0.90

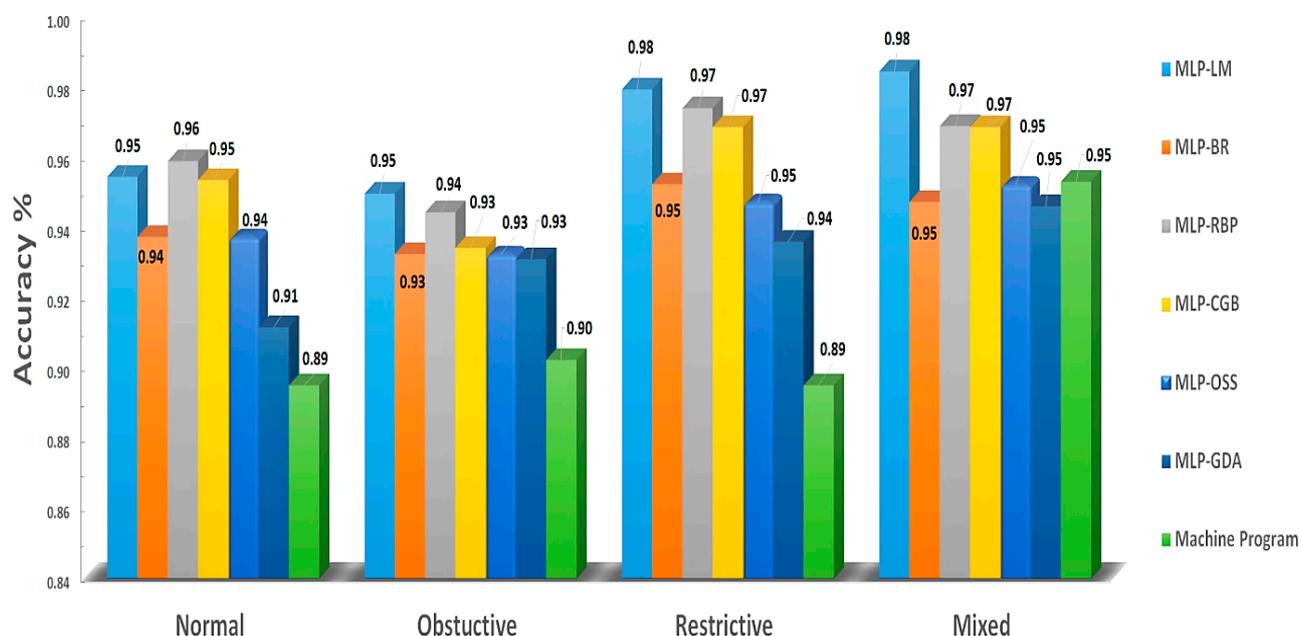


Figure 4. The accuracy of the PFT MLP medical decision support vs. the PFT Machine Program (i.e., IF/THEN Program). MLP—Multilayer Perceptron, LM—Levenberg Marquardt, BR—Bayesian Regularization, RBP—Resilient Back Propagation, CGB—Powell–Beale Conjugate Gradient, OSS—One Step Secant, GDA—Gradient Descent with Adaptive Learning Rule.

4. Results Discussion

Table 4 illustrates the MLP's classification accuracy of all 13 BP algorithms. The number of epochs was less than 60 for LM, BR, RBO, all CG, and OSS algorithms, whereas it was higher than 100 epochs for all GD, SOIT, and BT algorithms. The learning rate (LR) was equal to or less than 0.01 for all algorithms except BR, GDX, and GDA. The number of epochs was impossible for SOIT and BT because their structure does not utilize the stopping criteria or LR values. All MLP BP algorithms with the various architectures of numbers of neurons succeeded in yielding 92% to 99% accuracy on the training set and 87% to

92% accuracy on the test set. Therefore, these results sustain this paper's question statement (i.e., hypothesis) that the 23 respiratory parameters from ERS/ATS can successfully operate a PFT medical decision support system. This one is the first finding in this research paper.

Table 5 demonstrates the statistical performances of the preferred six BP algorithms on the test set. They resulted in 54 to 81% sensitivity, 92 to 94% specificity, 50 to 73% PPV, 91 to 93 NPV, and 88 to 94 AUC. On the one hand, these results revealed that the different BPs yielded different performance values. That is, a BP algorithm can outperform other BP algorithms in terms of one measure (e.g., accuracy), but it may underperform other BP algorithms (e.g., sensitivity). This result may be attributed to the low number of samples in the test set (50 PFT exams); this concept will be further discussed in the next paragraph. On the other hand, these results further support this paper's hypothesis that the 23 ERS/ATS standard respiratory parameters can be utilized satisfactorily in a PFT medical decision support system.

Figure 4 illustrates the second finding of this paper. It shows that the performance of an operating MLP BP algorithm is not necessarily the same for all outputs (decisions). For example, the LM resulted in approximately 95% accuracy on normal samples, while it yielded 95%, 98%, and 99% in classifying obstructive, restrictive, and mixed disorders, respectively. This marginal disparity occurred in all BP algorithms, the RB, BRP, CGB, OSS, and GDA. Consequently, as we mentioned in the previous section, a BP may exceed other BP algorithms in classifying one output (e.g., normal), but it may go beyond other BP algorithms for other decisions (e.g., restrictive disease). This means that the six BP algorithm outputs disperse slightly among each other. Therefore, the second finding in this paper is that different BP algorithms for the same type of ANN may yield different performances on different possible medical outputs (i.e., diagnostic decisions). This result can be applied to other medical decision support tasks other than the PFT exam.

Furthermore, Figure 4 reveals further findings. On the one hand, it indicates that the classification performances from the PFT machine program varied among different medical decisions, similarly to what was noticed from the MLP, as explained in the previous section. On the other hand, Figure 4 shows that all six MLP BP algorithms can outperform the built-in machine program in classifying the normal, obstructive, and restrictive conditions. However, only the LM, BR, and RBP exceeded the machine program in identifying the mixed condition. Therefore, the third finding in this paper is that by providing 23 ATS/ERS respiratory parameters as inputs to ANN, an MLP neural network could generate better performance than the PFT machine's diagnostic software. This finding was not similarly addressed in the literature. For instance, the LM algorithm exceeded the performance of the machine program by 6%, 5%, 9%, and 3% for classifying normal, obstructive, restrictive, and mixed cases, respectively. This result corresponds to a 6% average increment in classifying PFT cases. It is attributed to the capabilities of any ANN to analyze sophisticated nonlinear data; the matter is not feasible by an IF/THEN program.

Since the LM, BR, and RBP outperform the built-in machine program in classifying all PFT diagnosis decisions, including the mixed cases, they were utilized to plot the ROC curves, which is the comprehensive statistical parameter to observe any medical decision support capabilities. A ROC curve plots variations in sensitivity and specificity due to applying different threshold values in classifying outputs. Figure 5 shows the resulting ROC curves.

Table 6 shows results from some attempts presented in the literature. Regarding the number of samples, the report by Loachimescu et al. [15] utilized more than 15,000 PFT exams (i.e., patients). They utilized three respiratory features and a signal processing feature (i.e., the area under the expiratory flow–volume curve in Figure 1). They reported 83.5% and 91.6% accuracy, utilizing three and four features, respectively, to classify the same four respiratory cases (i.e., normal, obstructive, restrictive, and mixed cases). These results were the most reliable due to the extensive training and test sets. Another study was conducted by Badnjevic et al. [14], in which more than 5000 PFT exams were uti-

lized. However, they attempted to classify COPD from normal exams only. In addition, Topalovic et al. [4,39] trained their system on 1500 samples.

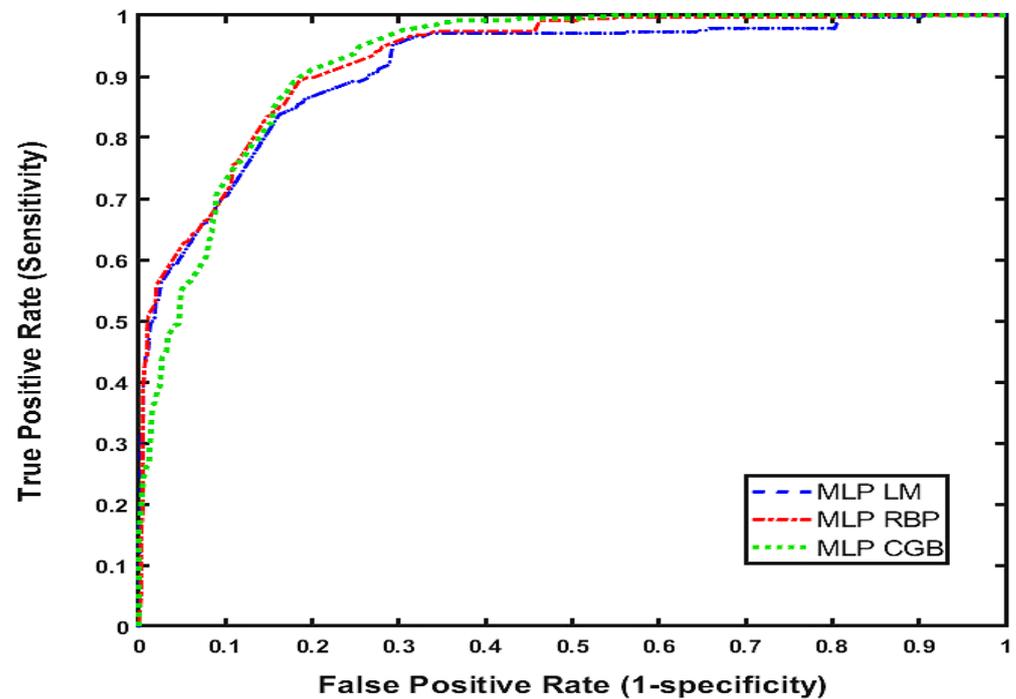


Figure 5. The MLP's ROC curves of LM, RBP, and CGB algorithms for all samples. MLP—Multilayer Perceptron, LM—Levenberg Marquardt, RBP—Resilient Back Propagation, CGB—Powell–Beale Conjugate Gradient.

The rest of the scientific papers reported in Table 6 used an almost similar number of samples as in this paper (i.e., 150 to 500 samples). They reported accuracies between 82 to 99% as the results in Table 4, but they explored part of the ERS/ATS parameters. Furthermore, they did not classify their samples into all four cases (four outputs) except for the work by Jafari et al. [12] and Topalovic et al. [4,39]. Jafari et al. utilized fitting coefficients of the processed signal and the four ERS/ATS parameters, whereas Topalovic et al. used about 21 PFT parameters and classified their samples into eight cases.

In terms of the type and number of input features (i.e., PFT parameters), most reports only explored the essential respiratory parameters (Group 1 in Table 1). However, except for the study made by Topalovic et al. [4,39], no reports had tested all 23 parameters, including the secondary respiratory parameters and the patient's anthropometric characteristics (Groups 2 and 3 in Table 1), as in this paper. Kalantary et al. [16] tested ten anthropometric parameters in the MLP decision support system (DSS) to discern normal from abnormal 130 PFT examinations. The results showed approximately 94% and 85% accuracies on the training and test sets, respectively. In this research work, all 23 features were included as inputs to MLP according to prior knowledge of these features' importance, which was confirmed by pulmonologists and international standards (ATS/ERS) for diagnosis [42]. However, specific selection methods of the unsupervised and supervised features could be implemented to investigate the significance of these features individually and combined [42,43]. While the unsupervised methods do not depend on the classification targets, and information loss might occur, the supervised methods suffer from multicollinearity and independency issues [42]. Therefore, feature selection methods were not utilized here to avoid any loss of valuable information, to accelerate the training process computationally, and to focus on MLP tuning parameters and algorithms.

Finally, in terms of the type of ANN technique, some researchers investigated MLP alone [11,12,14–16] or with other ANNs [4,7,13], as reported in Table 6. It seems that MLP is a successful ANN choice, since it yielded encouraging performances, comparable with

the results obtained in this experiment (Tables 4 and 5). However, other ANNs were also investigated, as in the references [4,6,10,21,22,24] reported in Table 6, leading to a range of accuracies as those obtained by the MLP neural network.

Table 6. Summary of specifications and accuracies from similar medical decision support systems of PFT exams.

Author/s	ANN Method	Features	#	Medical Decision	Samples	Accuracy [%]
Veezhinathan et al. [6]	RBF	FVC, FEV1, FEV1%, PEF, 3 pressures, 3 resistances	10	Normal and obstructive	100	90
Baemani et al. [11]	MLP	FVC, FEV1, FEV1%, PEF, FEF25–75%, age, height, weight, sex, smoker, race	11	Normal, obstructive, and restrictive	250	92.3
Manoharan et al. [7]	RBF/MLP	FVC, FEV1, FEV1%, PEF, FEF75%, 5 anthropometric, and 5 percentage values	15	Normal and abnormal	150	100/96
Sahin et al. [21]	SVM	FVC, FEV1, FEV1%	3	Normal, obstructive, and restrictive	499	97.3
Jafari et al. [12]	MLP (LM)	Predicted (FVC, FEV1, FEV1%, and PEF) + 6 fitted-curve coefficients	10	Normal, obstructive, restrictive, and mixed	205	97.6
Hakan et al. [17]	MLP	FVC, FEV1, FEV1%, FEF25-75, PEF	5	Normal, obstructive, and restrictive	486	98.7
Badnjevic et al. [13]	MLP (LM) + Fuzzy	FVC, FEV1, FEV1%, resistance, reactance, frequency (using IOS *)	6	Normal, COPD, and asthma	455	99.5
Spathis and Vlamos [22]	NN, NB, LogR, SVM, KNN, RFC	FEV1, FVC, FEV1%, PEF, MEF25/50/75/25-75, Sex, Smoke, pulse, O2 sat., age, and 9 symptoms	13	Asthma and COPD	132	89
Badnjevic et al. [14]	MLP	FVC, FEV1, FEV1%, VC, probability of disease	5	Normal, COPD, and Asthma	~5300	98.7
Topalovic et al. [4,39]	Decision Tree	FEV1, FVC, FEV1%, PEF, FEF25/50/75/25-75, Raw, sGaw, VC, RV, TGV, TLC, DLco, Kco, age, Smoke, CAT, gender, BMI	21	Asthma, COPD, OBD, NMD, TD, ILD, PVD, N	1430 + 50 + 136	82
Iadanza et al. [10]	RBNN + SVM + C5.0	FEV1/SVC, FEV1%, PEF, VC, TLC, RV, FRC, ERV, DLco, VA, DLco/VA, Height, Weight, Sex, Age	19	Mild, moderate, severe COPD	414	94.5
Loachimescu et al. [15]	MLP	Percent predicted (FVC, FEV1 & FEV1%) + sqrt AEX **	4	Normal, obstructive, restrictive, and mixed	15,308	83.5 91.6
Bodduluri et al. [24]	FCN + RFC	FEV1/FVC, FEV1 pred.	3	Normal, airway disease, emphysema, mixed	8980	80 Normal, 78 airway disease, 78 emphysema, 91 mixed
Kalantary et al. [16]	MLP DSS	Gender, age, weight, stature, body mass index, smoking, type of work, fat mass, fat free mass, and work history.	10	Normal and abnormal	130	93.6 (train) 84.6% (test) 91.5 (all)
This research work	MLP	23 parameters as specified by ATS and ERS (Table 1)	23	Normal, obstructive, restrictive, and mixed	201	92–99 training, 87–92 test

* IOS—Impulse Oscillometry System. ** AEX—Area Under the Expiratory Flow–Volume curve.

For example, Spathis and Vlamos [22] investigated machine learning techniques such as Naïve Bayes (NB), Logistic Regression (LogR), Neural Network (NN), Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Decision Tree (DT), and Random Forest

Classifier (RFC) to classify 132 PFT samples into two medical respiratory conditions, the COPD, and asthma. In contrast, Iadanza et al. [10] evaluated three models to design automatic learning techniques to classify 414 PFT samples into three levels of COPD severity, utilizing the Radial Basis Neural Network (RBNN), SVM, and C5.0, which is an automatic classifier function from the IBM SPSS Modeler capable of selecting the most suitable classification algorithm for any given data. Bodduluri et al. [24] attempted to use a deep-learning model (i.e., Fully Convolutional Network (FCN)) and RFC, employing PFT flow–volume curves of nearly 9000 samples. The latter three studies [10,22,24] recorded 78–94.5% accuracies, which overlap with records utilizing MLPs [7,11–16] in Table 6, alongside the results obtained in this paper (as reported in Table 4).

It is worth mentioning that the number of classes to which the artificial intelligence (AI) technique will classify is determinant to the complexity of the modeled relationship input–output classes. Topalovic et al. [4,39] were the first to classify their samples into eight different cases (categories), making the classification task for machine learning even harder to achieve. Those eight cases are asthma, COPD, other obstructive diseases (OBD), neuromuscular disorder (NMD), thoracic deformity (TD), interstitial lung disease (ILD), pulmonary vascular disease (PVD), and normal condition.

As observed from the specific eight cases (which contain the four previous ones used in our study) the complexity to distinguish between these classes is relatively higher than that of fewer classes (e.g., four cases). That is why the classification accuracy reported in this study was relatively lower than that of other studies with fewer output classes (≤ 4). Furthermore, Topalovic et al. used very extensive examinations, such as the diffusing capacity of the lung for carbon monoxide, to obtain more parameters and support the pulmonologist's decision.

It is also of importance to highlight that the reports in Table 6 have differences in the terms of the ANN specifications (i.e., ANN architecture). Possible specifications, but not limited to, are the number of hidden layers, number of neurons, type of activation function, and the normalization scale of the input data. The distribution of the PFT samples, whether they cover the entire input space or not, also has an important role in enriching the learning process. They all affect the performance of any AI medical decision system. These AI specifications were not reflected in comparison with the literature because this is beyond the purpose of this research work.

However, it is important to indicate that other MLP architecture specifications may further optimize or refine the results; one of which is the assessment of different activation functions that govern ANN outputs. Another factor is the repetition of the experiment using several iterations of the random 75–25% splitting protocol of the dataset, which would ascertain ANN's performance. Increasing the number of hidden layers is another impact in ANN science; this might improve the classification performance, but it would also increase the training time and the computer capacity requirements. To end, increasing the number of PFT samples, as in reports by Loachimescu et al. [15] and Badnjevic et al. [14], would finalize the merits of using the 23 ERS/ATS respiratory parameters as inputs to the PFT medical decision support system.

5. Conclusions

This paper demonstrated that the 23 ATS/ERS international standard PFT parameters could be used as input for a reliable medical decision support system. A system was proposed using the MLP neural network technique with 13 BP algorithms. It yielded 92–99% and 87–92% accuracy on training and test sets from 201 PFT samples. On the one hand, the results outperformed the PFT built-in diagnostic software. On the other hand, they overlapped with records by other researchers who utilized part of the ATS/ERS parameters. In addition, a slight disparity in accuracy was observed among the BP algorithms for each medical output, indicating that it may be worthwhile to utilize the best BP algorithm for each medical decision.

Author Contributions: All authors contributed to the study conception and design. Conceptualization, A.A.A. and A.-R.A.-H.; methodology, A.A.A. and R.D.F.; software, A.A.A. and R.D.F.; validation, A.-R.A.-H. and P.V.; formal analysis, A.A.A. and A.-R.A.-H.; investigation, A.A.A. and P.V.; data curation, A.A.A. and R.D.F.; writing—original draft preparation, A.A.A., A.-R.A.-H. and P.V.; writing—review and editing, A.A.A., A.-R.A.-H., R.D.F. and P.V.; supervision, A.-R.A.-H. and P.V.; funding acquisition, P.V. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data of our study are available upon request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Peters, J.I.; Levine, S.M. Introduction to Pulmonary Function Testing (Chapter 27). In *Respiratory Disorders*; McGraw Hill Medical: New York, NY, USA, 2009; pp. 271–278.
2. Miller, M.R.; Hankinson, J.; Brusasco, V.; Burgos, F.; Casaburi, R.; Coates, A.; Crapo, R.; Enright, P.; Van Der Grinten, C.P.M.; Gustafsson, P.; et al. Standardization of spirometry. *Eur. Respir. J.* **2005**, *26*, 319–338. [[CrossRef](#)] [[PubMed](#)]
3. Giri, P.C.; Chowdhury, A.M.; Bedoya, A.; Chen, H.; Lee, H.S.; Lee, P.; Henriquez, C.; MacIntyre, N.R.; Huang, Y.-C.T. Application of Machine Learning in Pulmonary Function Assessment Where Are We Now and Where Are We Going? *Front. Physiol.* **2021**, *12*, 678540. [[CrossRef](#)] [[PubMed](#)]
4. Topalovic, M.; Laval, S.; Aerts, J.-M.; Troosters, T.; Decramer, M.; Janssens, W. Automated Interpretation of Pulmonary Function Tests in Adults with Respiratory Complaints. *Respiration* **2017**, *93*, 170–178. [[CrossRef](#)] [[PubMed](#)]
5. Haykin, S. *Neural Networks and Learning Machines*; Prentice Hall: New York, NY, USA, 2009.
6. Veezhinathan, M.; Ramakrishnan, S. Detection of Obstructive Respiratory Abnormality Using Flow–Volume Spirometry and Radial Basis Function Neural Networks. *J. Med. Syst.* **2007**, *31*, 461–465. [[CrossRef](#)]
7. Manoharan, S.C.; Veezhinathan, M.; Ramakrishnan, S. Comparison of Two ANN Methods for Classification of Spirometer. *Data Meas. Sci.* **2008**, *8*, 53–57. [[CrossRef](#)]
8. Manoharan, S.C.; Ramakrishnan, S. Prediction of Forced Expiratory Volume in Pulmonary Function Test using Radial Basis Neural Networks and k-means Clustering. *J. Med. Syst.* **2009**, *33*, 347–351. [[CrossRef](#)]
9. Waghmare, K.A.; Wakode, B.V.; Chatur, P.N. Spirometry Data Analysis and Classification Using Artificial Neural Network: An Approach. *Int. J. Emerg. Technol. Adv. Eng.* **2012**, *2*, 67–70.
10. Iadanza, E.; Mudura, V.; Melillo, P.; Gherardelli, M. An automatic system supporting clinical decision for chronic obstructive pulmonary disease. *Health Technol.* **2020**, *10*, 487–498. [[CrossRef](#)]
11. Baemani, M.; Monadjemi, A.; Moallem, P. Detection of Respiratory Abnormalities Using Artificial Neural Networks. *J. Comput. Syst. Sci.* **2008**, *4*, 663–667. [[CrossRef](#)]
12. Jafari, H.; Arabalibeik, H.; Agin, K. Classification of Normal and Abnormal Respiration Patterns Using Flow Volume Curve and Neural Network. In Proceedings of the IEEE International Symposium on Health Informatics and Bioinformatics (HIBIT), Antalya, Turkey, 20–22 April 2010. [[CrossRef](#)]
13. Badnjevic, A.; Cifrek, M.; Koruga, D.; Osmankovic, D. Neuro-fuzzy classification of asthma and chronic obstructive pulmonary disease. *BMC Med. Inform. Decis. Mak.* **2015**, *15*, S1. [[CrossRef](#)]
14. Badnjevic, A.; Gurbeta, L.; Custovic, E. An Expert Diagnostic System to Automatically Identify Asthma and Chronic Obstructive Pulmonary Disease in Clinical Settings. *Sci. Rep.* **2018**, *8*, 11645. [[CrossRef](#)] [[PubMed](#)]
15. Loachimescu, O.C.; Stoller, J.K. An Alternative Spirometric Measurement: Area under the Expiratory Flow–Volume Curve. *Ann. Am. Thorac. Soc.* **2020**, *17*, 582–588. [[CrossRef](#)] [[PubMed](#)]
16. Kalantary, S.; Pourbabaki, R.; Jahani, A.; Yarandi, M.S.; Samiei, S.; Jahani, R. Development of a decision support system tool to predict the pulmonary function using artificial neural network approach. *Concurr. Comput. Pract. Exper.* **2021**, *33*, e6258. [[CrossRef](#)]
17. Hakan, A.; Guler, I.; Sener, M.U. A Low-Cost Mobile Adaptive Tracking System for Chronic Pulmonary Patients in Home Environment. *Telemed. e-Health* **2013**, *19*, 24–30. [[CrossRef](#)]
18. Trivedy, S.; Goyal, M.; Mohapatra, P.R.; Mukherjee, A. Design and Development of Smartphone Enabled Spirometer with a Disease Classification System Using Convolutional Neural Network. *IEEE Trans. Instrum. Meas.* **2020**, *69*, 7125–7135. [[CrossRef](#)]
19. Er, O.; Yumusak, N.; Temurtas, F. Chest diseases diagnosis using artificial neural networks. *Expert Syst. Appl.* **2010**, *37*, 7648–7655. [[CrossRef](#)]
20. Kavitha, A.; Sujatha, C.M.; Ramakrishnan, S. Prediction of Spirometric Forced Expiratory Volume (FEV1) Data Using Support Vector Regression. *Meas. Sci. Rev.* **2010**, *10*, 63–67. [[CrossRef](#)]

21. Sahin, D.; Ubeyli, E.; Ilbay, G. Diagnosis of Airway Obstructive or Restrictive Spirometry Patterns by Multiclass Support Vector Machines. *J. Med. Syst.* **2010**, *34*, 967–973. [CrossRef]
22. Spathis, D.; Vlamos, P. Diagnosing asthma and chronic obstructive pulmonary disease with machine learning. *Health Inform. J.* **2009**, *25*, 811–827. [CrossRef]
23. Junwale, P.D.; Bhade, A.W.; Chatur, P.N. Statistical Data Mining Approach for Spiro metric Data Classification: Review Paper. *Int. J. Comput.* **2012**, *3*, 348–351.
24. Bodduluri, S.; Nakhmani, A.; Reinhardt, J.M.; Wilson, C.G.; McDonald, M.-L.; Rudraraju, R.; Jaeger, B.; Bhakta, N.R.; Castaldi, P.J.; Sciruba, F.C.; et al. Deep neural network analyses of spirometry for structural phenotyping of chronic obstructive pulmonary disease. *JCI Insight* **2020**, *5*, e132781. [CrossRef] [PubMed]
25. Murdaca, G.; Caprioli, S.; Tonacci, A.; Billeci, L.; Greco, M.; Negrini, S.; Cittadini, G.; Zentilin, P.; Spagnolo, E.V.; Gangemi, S. A Machine Learning Application to Predict Early Lung Involvement in Scleroderma: A Feasibility Evaluation. *Diagnostics* **2021**, *11*, 1880. [CrossRef] [PubMed]
26. Kavitha, A.; Sujatha, C.M.; Ramakrishnan, S. Evaluation of Flow–Volume Spirometric Test Using Neural Network Based Prediction and Principal Component. *J. Med. Syst.* **2011**, *35*, 127–133. [CrossRef] [PubMed]
27. Pellegrino, R.; Viegi, G.; Brusasco, V.; Crapo, R.O.; Burgos, F.; Casaburi, R.; Coates, A.; Van Der Grinten, C.P.M.; Gustafsson, P.; Hankinson, J.; et al. Interpretative strategies for lung function tests. *Eur. Respir. J.* **2005**, *26*, 948–968. [CrossRef] [PubMed]
28. World Health Organization (WHO). The Top 10 Causes of Death. Available online: <https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death> (accessed on 6 July 2021).
29. Al-Naami, B.; Fraihat, H.; Al-Nabulsi, J.; Gharaibeh, N.Y.; Visconti, P.; Al-Hinnawi, A.R. Assessment of Dual Tree Complex Wavelet Transform to improve SNR in collaboration with Neuro-Fuzzy System for Heart Sound Identification. *Electronics* **2022**, *11*, 938. [CrossRef]
30. Tseng, H.-J.; Veeraraghavan, S.; Henry, T.S.; Mittal, P.K.; Little, B.P. Pulmonary Function Tests for the Radiologist. *Radio Graph.* **2017**, *37*, 1037–1058. [CrossRef]
31. COSMED Company for Pulmonary Function Equipment. Quark PFT—Innovative Modularity and Networking for Truly Customised PFT Solutions. Available online: <https://www.cosmed.com/en/products/pulmonary-function/quark-pft> (accessed on 6 July 2021).
32. Demuth, H.; Beale, M.; Hagan, M. *Neural Network Toolbox 6: User’s Guide*; MathWorks Inc.: Natick, MA, USA, 2017; Available online: <http://www.mathworks.com> (accessed on 14 June 2021).
33. Stathakis, D. How many hidden layers and nodes? *Int. J. Remote Sens.* **2009**, *30*, 2133–2147. [CrossRef]
34. Kurkova, V. Kolmogorov’s theorem and multilayer neural networks. *Neural Netw.* **1992**, *5*, 501–506. [CrossRef]
35. Huang, G.B. Learning capability and storage capacity of two-hidden-layer feed forward networks. *IEEE Trans. Neural Netw.* **2003**, *14*, 274–281. [CrossRef]
36. Hecht-Nielsen, R. Kolmogorov’s mapping neural network existence theorem. In Proceedings of the IEEE First Annual International Conference on Neural Networks, San Diego, CA, USA, 21–24 June 1987; Volume III, pp. 11–14.
37. Zhu, W.; Zeng, N.; Wang, N. Sensitivity, specificity, accuracy, associated confidence interval and ROC analysis with practical SAS implementations, Maryland, Baltimore. In Proceedings of the Health Care and Life Sciences NESUG 2010, Baltimore, MD, USA, 14–17 November 2010; pp. 1–9. Available online: <https://www.lexjansen.com/nesug/nesug10/hl/hl07.pdf> (accessed on 2 September 2020).
38. Al-Naami, B.; Fraihat, H.; Abu Owida, H.; Al-Hamad, K.; De Fazio, R.; Visconti, P. Automated Detection of Left Bundle Branch Block from ECG signal utilizing the Maximal Overlap Discrete Wavelet Transform with ANFIS. *Computers* **2022**, *11*, 93. [CrossRef]
39. Topalovic, M.; Das, N.; Burgel, P.-R.; Daenen, M.; Derom, E.; Haenebalcke, C.; Janssen, R.; Kerstjens, H.A.; Liistro, G.; Louis, R.; et al. Artificial intelligence outperforms pulmonologists in the interpretation of pulmonary function tests. *Eur. Respir. J.* **2019**, *53*, 1801660. [CrossRef] [PubMed]
40. Hoens, T.R.; Chawla, N.V. Imbalanced Datasets: From Sampling to Classifiers. In *Imbalanced Learning: Foundations, Algorithms, and Applications—Chapter 3*; He, H., Ma, Y., Eds.; Wiley-IEEE Press: Hoboken, NJ, USA, 2013; p. 216. [CrossRef]
41. Kuhn, M.; Johnson, K. Remedies for Severe Class Imbalance. In *Applied Predictive Modeling*; Springer Science & Business Media: New York, NY, USA, 2013; pp. 419–443. [CrossRef]
42. Guyon, I.; Elisseeff, A. An Introduction to Variable and Feature Selection. *J. Mach. Learn. Res.* **2003**, *3*, 1157–1182. [CrossRef]
43. Kuhn, M.; Johnson, K. An Introduction to Feature Selection. In *Applied Predictive Modeling*; Springer Science & Business Media: New York, NY, USA, 2013; pp. 487–519. [CrossRef]