

# SCIENTIFIC REPORTS



OPEN

## An Expert Diagnostic System to Automatically Identify Asthma and Chronic Obstructive Pulmonary Disease in Clinical Settings

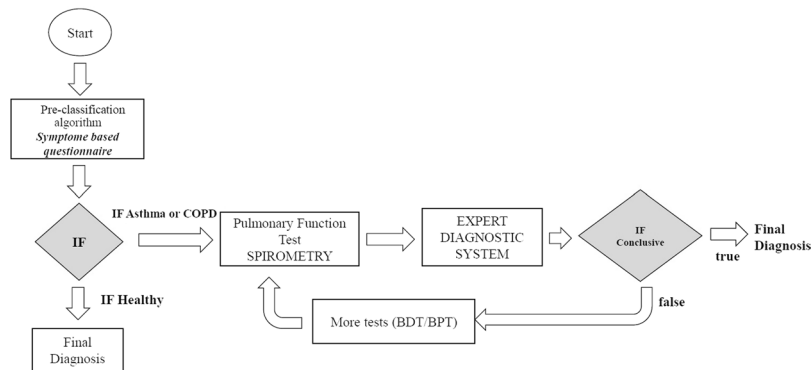
Almir Badnjevic<sup>1,3,4</sup>, Lejla Gurbeta<sup>1,2,3</sup> & Eddie Custovic<sup>5</sup>

Respiratory diseases such as asthma and chronic obstructive pulmonary disease (COPD), are affecting a huge percentage of the world's population with mortality rates exceeding those of lung cancer and breast cancer combined. The major challenge is the number of patients who are incorrectly diagnosed. To address this, we developed an expert diagnostic system that can differentiate among patients with asthma, COPD or a normal lung function based on measurements of lung function and information about patient's symptoms. To develop accurate classification algorithms, data from 3657 patients were used and then independently verified using data from 1650 patients collected over a period of two years. Our results demonstrate that the expert diagnostic system can correctly identify patients with asthma and COPD with sensitivity of 96.45% and specificity of 98.71%. Additionally, 98.71% of the patients with a normal lung function were correctly classified, which contributed to a 49.23% decrease in demand for conducting additional tests, therefore decreasing financial cost.

CHRONIC obstructive pulmonary disease (COPD) is a chronic inflammatory lung disease that causes obstructed airflow in the lungs<sup>1,2</sup>. Up to 75% of all COPD patients are not diagnosed. Currently, the COPD mortality rate exceeds that of lung cancer and breast cancer combined<sup>3-6</sup>, as 200,000 to 300,000 deaths in Europe alone are COPD related. Similarly, asthma is a chronic inflammatory impairment of airways, which, as a result, becomes hyperactive and generates increased mucus, mucosal swelling and contraction of smooth airway muscles. These factors all contribute to airway obstruction. With respect to other chronic respiratory diseases, asthma has relatively low fatality rate, but the prevalence of asthma as well as the costs of asthma treatment and care has increased in recent decades<sup>1,2</sup>.

A major challenge in chronic disease management, especially in non-specialized clinics, is the number of patients with chronic respiratory diseases, such as asthma or COPD, who are either inaccurately diagnosed or misdiagnosed for having other respiratory diseases such as the common cold, acute bronchitis or pneumonia<sup>6,7</sup>. Over the years, various evidence-based guidelines for the prevention, diagnosis and management of chronic respiratory diseases have been developed to assist medical professionals. The Global Initiative for Chronic Obstructive Lung Disease (GOLD)<sup>8</sup> and Global Initiative for Asthma (GINA)<sup>9</sup> has published guidelines for medical professionals based on their latest research and recommendations. Despite the availability of these guidelines, the lack of knowledge among non-specialized medical professionals is still a leading issue in the correct diagnosis of these respiratory diseases<sup>10</sup>. In 2008, Yawn and Wollan<sup>10</sup> showed that many primary medical professionals are unable to distinguish asthma from COPD and do not recognize that women are at a higher risk for COPD than men. When COPD symptoms are misdiagnosed as asthma, women receive the wrong treatment, and the correct COPD treatment is delayed. This misdiagnosis has serious adverse consequences in terms of disease burden and risk of future exacerbations.

<sup>1</sup>International Burch University, Faculty of Engineering and Natural Sciences, Genetics and Bioengineering Department, Sarajevo, Bosnia and Herzegovina. <sup>2</sup>Technical faculty University of Bihac, Bihac, Bosnia and Herzegovina. <sup>3</sup>Medical Device Inspection Laboratory Verlab Ltd, Sarajevo, Bosnia and Herzegovina. <sup>4</sup>Faculty of Electrical Engineering University of Sarajevo, Sarajevo, Bosnia and Herzegovina. <sup>5</sup>La Trobe Innovation & amp, Entrepreneurship Foundry at La Trobe University Melbourne, Melbourne, Australia. Correspondence and requests for materials should be addressed to A.B. (email: [almir.badnjevic@ibu.edu.ba](mailto:almir.badnjevic@ibu.edu.ba))



**Figure 1.** A block diagram of the entire expert diagnostic system.

The usage of computer-based methods in medical diagnoses are on the rise and are gradually improving the quality of medical services by utilizing larger datasets of symptoms and patient history, as well as diagnostic test results for diagnosis. Artificial intelligence and machine learning methods are being used to create expert systems that can employ human knowledge and solve problems that ordinarily require direct human expertise<sup>11,12</sup>. Beginning in the 1990s and increasing in the 2000s, expert systems based on machine learning methods, such as artificial neural networks (ANNs) and fuzzy logic (FL) were used for the detection of different types of diseases, including respiratory diseases. Walia *et al.*<sup>13</sup> presented a systematic approach for design and identification of tuberculosis using a fuzzy based decision support system. Asaithambi *et al.*<sup>14</sup> classified respiratory abnormalities using an adaptive ANN and FL inference system based only on spirometry (SPIR) or impulse oscillometry measurements (IOS). Stavrakoudis *et al.*<sup>15</sup> separated lung sounds, obtained from patients with pulmonary pathology using a recurrent neuro-fuzzy filter. Several studies focusing on the use of different types of ANN architectures for classification of respiratory diseases with high classification accuracies developed on various datasets have been undertaken<sup>16–20</sup>. Other machine learning techniques, such as Random Forests, Gradient Boosting, or even Logistic Regression can also be used for prediction and disease classification<sup>17–19</sup>, but neural networks show optimal performance for this task when there is larger number of samples in databases. However, earlier studies based their classification efforts predominantly on SPIR and/or IOS measured test results, i.e. static assessment of patients. Even though high classification accuracies were reported, additional full dynamic assessments of patients are needed to acquire knowledge about a patient's symptoms and pulmonary function. Measurement of lung function is important because patients often do not recognize mild symptoms or do not attach importance to them, especially if they persist for a longer period of time. Furthermore, no previously developed system was tested in real-time in clinical settings, that is, all systems were developed and validated offline.

The purpose of this study was to evaluate the impact of introducing Expert Diagnostic System (EDS) into a healthcare system. Our hypothesis was that an accurate EDS could differentiate patients with asthma, COPD and a normal lung function with a classification rate of over 90%. To test our hypothesis, we developed an EDS based on data from 3657 patients and utilized combined ANN and FL algorithms. We then implemented the developed system in a local hospital in Sarajevo, Bosnia and Herzegovina. The EDS system was tested on 1650 patients, and our results showed that our system could achieve a classification rate well above 90%. Our results demonstrate that the developed EDS system is reliable and such an automated diagnosis tool would be beneficial for healthcare institutions, especially in primary care and remote healthcare institutions.

## Materials and Methods

**Implemented expert diagnostic system.** A block diagram of the EDS for automated identification of COPD and asthma is presented in Fig. 1. It consists of (1) a pre-classification algorithm used to determine whether confirmatory respiratory function tests are needed based on a symptom questionnaire and (2) a classifier based on a combination of a single-layer ANN and FL.

The first step of this EDS is to utilize a pre-classification algorithm to perform initial screening of all patients (healthy vs. COPD/asthma) based on answers provided to a symptom-based questionnaire. From the patient's answers, the probability of having COPD or asthma is determined. If the calculated probability indicates that the patient falls into the "Normal" category, the patient is diagnosed as "Healthy" and the process ends. However, if the pre-classification algorithm indicates that the probability of being affected by asthma or COPD is greater than 50%, SPIR tests need to be performed. The results of the pulmonary function test are then used as inputs for the implemented Expert Diagnostic System (EDS). The EDS determines whether a conclusive diagnosis, based on spirometry measurements, can be derived (i.e. resulted in a final diagnosis) or further tests (such as bronchodilation testing - BDT and/or broncho-provocation testing - BPT) are required to reach a more robust decision. Once all required tests are analyzed by the EDS, the final decision is obtained.

**Pre-classification algorithm.** To determine the need of running the EDS and performing confirmatory respiratory tests on patients, a pre-classification algorithm was developed. Inputs to the pre-classification algorithms are patient's answers on seven "yes" or "no" questions about symptoms as shown in Table 1. Based on these answers, the algorithm then calculates the probability of the patient being healthy or diseased (asthma, COPD). These calculations make it possible to identify patients that are required to take a confirmatory test for diagnosis

No.	Question	Answer	Asthma (As)	COPD (C)
Q <sub>1</sub>	Does the patient exceed the age of 40 years?	If answered affirmatively the patient <b>may</b> have COPD. If the answer is negative the patient <b>may have asthma</b> .	sf <sub>1As</sub>	sf <sub>1C</sub>
Q <sub>2</sub>	Does the patient experience problems while exercising or performing low intensity activities?	If answered affirmatively the patient <b>may</b> have COPD.	sf <sub>2As</sub>	sf <sub>2C</sub>
Q <sub>3</sub>	Does the patient cough at night or right after waking up?	If answered affirmatively the patient <b>may</b> have <b>asthma</b> or <b>heart disease</b> .	sf <sub>3As</sub>	sf <sub>3C</sub>
Q <sub>4</sub>	Does the patient suffer from an abundant presence of mucus in the throat?	If answered affirmatively the patient <b>may</b> have COPD.	sf <sub>4As</sub>	sf <sub>4C</sub>
Q <sub>5</sub>	Does the patient experience high-pitched breath sounds during the morning?	If answered affirmatively the patient <b>may</b> have COPD.	sf <sub>5As</sub>	sf <sub>5C</sub>
Q <sub>6</sub>	Does the patient experience high pitched breath sounds at night, or while working out, or while performing low intensity activities?	If answered affirmatively the patient <b>may</b> have <b>asthma</b> .	sf <sub>6As</sub>	sf <sub>6C</sub>
Q <sub>7</sub>	Does the patient experience a choking sensation, while at rest?	If answered affirmatively the patient <b>may</b> have COPD or <b>asthma</b> .	sf <sub>7As</sub>	sf <sub>7C</sub>

**Table 1.** questionnaire form and significance factors for each question related to asthma and copd. \*sf<sub>jk</sub> for the *j*<sup>th</sup> question and *k*<sup>th</sup> disease Given the abovementioned parameters, the probability (p<sub>k</sub> for the *k*th disease) of the presence of asthma or COPD in a patient can be calculated using the following equation.

and determine the most appropriate test for the patient. The seven questions, shown in Table 1, were formed in accordance to GINA and GOLD guidelines and took into consideration factors such as age, history of exposure to tobacco and/or other smokes at an age over 40, history of progressive, exertion or persistent dyspnea, occupational dusts, chronic intermittent or non-productive cough and chronic sputum.

To quantify the answers given by patients, each answer from the questionnaire was given an option factor (O<sub>j</sub> for the *j*<sup>th</sup> question). A binary value was assigned to O<sub>j</sub>; with a positive answer assigned as 1 and a negative answer as 0. To calculate the probability of disease, a weight was assigned to each answer according to Table 1. These weights were determined in accordance to GINA and GOLD guidelines since not all symptoms are equally significant in the diagnosis of COPD or asthma.

To calculate the probability of disease, a constant of normalization (A<sub>k</sub> for the *k*th disease) for asthma and COPD was calculated as the inverse of the sum of the factors:

$$A_k = \frac{1}{\sum_{i=1}^{i=N} sf_{jk}} \quad (1)$$

$$p_k = A_k \times \sum_{i=1}^{i=N} (sf_{jk} \times O_j) \times 100(\%) \quad (2)$$

The probabilities obtained in this process represent the certainty that a patient has indicated type of respiratory disease (asthma or COPD).

Once the patient was pre-classified, it was determined whether a confirmatory test was necessary and, if so, which test was to be performed. The information gathered from these tests was later used as input data for the Expert System. Diagnoses, established by medical professionals, were used to evaluate the pre-classification algorithm and improve it. Both the error (E<sub>k</sub> for the *k*th disease) and the accuracy of the pre-classification could be obtained once these variables were known by using equations (3) and (4).

$$E_k = P_k^c - p_k \quad (3)$$

where p<sub>k</sub> is the diagnosis performed by the medical professional.

$$Accuracy = \frac{p_k}{P_k^c} \times 100(\%) \quad (4)$$

**The expert diagnostic system (ANN-FL logic).** In this paper, a feedforward neural network was used for the classification of COPD and asthma since it is proven that this type of neural network architecture is sufficient for these tasks<sup>21–25</sup>. The results from the pulmonary function test, parameters: VC (Vital Capacity), FEV1 (Forced Expiratory Volume in the 1<sup>st</sup> second), FVC (Forced Vital Capacity) and FEV1% (Tiffeneau index), were introduced as inputs in the ANN, resulting in 4 neurons in the input layer. The fifth input is the result of the pre-classification algorithm, which represents the probability of disease. To standardize the range of independent variables all inputs to neural network were normalized using min-max normalization. This neural network architecture has one output neuron that is the index of 10 output classes being as follows: ASTHMA1, ASTHMA2, ASTHMA3, ASTHMA4, GOLD1, GOLD2, GOLD3, GOLD4, NORMAL and INCONCLUSIVE.

Each neuron in the hidden layer performs a weighted summation of the inputs, which is then passed to a non-linear activation function. The hidden layer neurons use nonlinear hyperbolic tangent sigmoid transfer function, while the linear activation function was used for the output layer. Although this function is commonly used

in regression problems<sup>26</sup>, the system with these transfer functions in each layer showed optimum performance. A weight matrix ( $I_{ws}$ , where  $s = 1 \dots S$  (input neurons) and  $n = 1 \dots N$  (hidden neurons)) was automatically defined to the links connecting input and hidden neurons ( $H_n$ ) following the training algorithm. An activation function ( $H_{On}$ ) was computed for each hidden neuron, taking into account its bias ( $H_{bn}$ ):

At this point, a new weight matrix ( $H_{w_{n,r}}$  where  $n = 1 \dots N$  (hidden neurons) and  $r = 1 \dots R$  (output neurons)) was assigned to the links, connecting the neurons in the hidden and the output layer ( $O_r$ ). The outputs ( $Y_r$ ) were then calculated by considering the abovementioned parameters and the corresponding bias in each output neuron ( $O_{br}$ ):

The number of neurons in the hidden layer was determined by evaluating the performance of the neural network architectures with 5, 10, 14 and 20 neurons. The tested neural network architectures were compared based on accuracy of classification of COPD and asthma.

The fuzzy logic output classifier was developed based on recommendations by experts in this field<sup>25–30</sup> and international guidelines, GINA and GOLD. Input variables for the implemented fuzzy logic output classifier are the result of BDT/BPT tests and information about the disease probability. The fuzzy logic rules were defined based on the severity of the disease, asthma or COPD and information about disease probability. Based on the fuzzy rules, outputs for the diagnosis were defined as COPD, asthma or healthy<sup>20</sup>.

**Algorithm training and testing.** During the development of a neural network, instead of dividing the dataset into distinct groups for training, validation and testing with a fixed number of samples,  $k$ -fold cross validation was used. The overall dataset for the development of the neural network was divided into  $k$  subsets. At each learning iteration, the neural network was trained on  $k - 1$  subsets, then tested on the one subset, which was not used during training. This process is repeated  $k$  times, each time using a different test set chosen from the  $k$  available divisions of the training data, until all possible test sets have been used. The  $k$  test set performances for each model are averaged, and the model with the highest average performance is chosen as the one most likely to perform well on unseen data. For training, the Levenberg – Marquardt algorithm (LMA), which is a common training algorithm in data classification was used<sup>29</sup>. As a measure of performance, at each training iteration, the Mean Square Error (MSE) between the predicted and actual values was calculated, where  $n$  is total number of samples:

$$MSE = \frac{1}{n} \sum_{i=1}^n (X_{predicted} - X_{actual})^2 \quad (5)$$

The classification accuracy in relation to the number of neurons in the hidden layer was also examined. While the number of input and output neurons was determined with data structure and process modeling, the performance of training was dependent on the complexity of the neural network and the number of neurons in the hidden layer<sup>31</sup>. A poorly defined number of neurons in the hidden layer can result in overfitting the problem. This overfitting leads to very high training performance (accuracy of classification >90%) and very poor testing performance (accuracy of classification <60%). There are various methods for choosing the fixed number of neurons in the hidden layer, but there is no generally accepted method for determining the number of neurons in a single hidden layer that would efficiently approximate any given function or process. Despite the new methods developed for this purpose, most researchers use a trial rule<sup>25</sup>. This rule was used in this study as well. The number of neurons in the hidden layer was set to be 5, 12, 14, 17 and 20 and their performance was calculated. The smallest MSE was achieved with 17 neurons in the hidden layer (12.485), but since the neural network with 14 neurons in the hidden layer (12.569) had a marginally poorer performance, 14 neurons were chosen for further development to avoid a more complex architecture.

**Sample collection.** Following internationally accepted medical practices for diagnosis of COPD and asthma, a prospective study was developed to generate the dataset for design, validation and real-time use of EDS, for automated identification of asthma and COPD. Before starting the study, the ethics board approval for human subject testing from the Hospital Sarajevo was obtained, as well as the patients' informed consent. Healthcare institutions also approved all methods and procedures which were performed in accordance with the relevant guidelines and regulations. Finally, a dataset of 3657 samples based on patient reports was generated, which included relevant information established by medical professionals. Out of the 3657 samples, 3000 previously collected 2014 and 2015 samples were used for development and training of ANN, while 657 samples were used for validation of fuzzy classifier in EDS.

The whole implemented EDS was tested with 1650 samples in real-time at the Pulmonary Clinic of Sarajevo and primary healthcare institutions during 2016. The class distribution of the dataset is presented in Table 2. In the dataset of 1650 samples, 1495 samples were samples of diseased subjects (859 of asthma and 636 of COPD) and 155 were samples of healthy subjects. During the real-time clinical testing of the EDS, parallel diagnoses were performed by medical professional and the EDS. The EDS diagnosis was revealed to medical professionals after they formalized the diagnosis for the same patient. At the end, the EDS diagnosis was finally compared with the medical professional diagnosis.

Each sample contained seven features representing the patient symptoms information and four features representing the results of the SPIR tests. Additionally, some samples had features representing the results of BDT/BPT test in the case that it was undertaken. Only patients with confirmed a diagnosis were subject to this study. Diagnoses were performed by medical professionals following clinical assessment according to international guidelines<sup>8,9</sup>. Baseline assessments consisted of screening for patient symptoms using symptom based questionnaires' or interview conducted by a medical professional. All spirometry lung function tests were obtained using the CareFusion "Master Screen" device, which measured, derived and calculated all the required spirometry parameters that were features of samples.

	Asthma	COPD	Healthy	
Total number of samples	5307	2485	1514	1308
Samples used for ANN development *2014–2015	3000 (56.53%)	1267 (50.98%)	689 (45.51%)	1044 (79.82%)
Samples used for fuzzy classifier development *2014–2015	657 (12.38%)	359 (14.45%)	189 (12.48%)	109 (8.33%)
Real – time testing in Pulmonary Clinic of Sarajevo *during 2016	1650 (31.09%)	859 (34.57%)	636 (42.01%)	155 (11.85%)
<b>*Etymology of real – time testing data</b>				
Sex	896 Male		754 Female	
Age (Average)	44.12		49.23	
Symptoms	72.3% of subjects have 3 or more symptoms of disease		81.6% of subjects have 3 or more symptoms of disease	
Other diseases diagnosed	15.7%		13.1%	
Consummation of medications during past 7 days	54.2%		49.7%	

**Table 2.** Dataset Distribution per Classes.

The etymology of the data (patient age, sex, treatment) used for development and training of EDS was not analyzed since they are part of already established disease diagnosis, while etymology of data from real-time testing of EDS was analyzed and presented in Table 2.

## Results

The pre-classification algorithm was validated on 2735 samples from the overall dataset. During validation of the algorithm, the accuracy of 95.17% for samples with COPD or asthma and 98.7% for pre-classification of healthy subjects were both achieved.

It should be noted that the results of the pre-classification algorithm are also checked (confirmed or discarded) by medical professional so the quality of patient care is not decreased. The rate of misdiagnosis by pre-classification algorithm is still lower than currently available information on COPD and asthma diagnosis.

By introducing computer-aided systems with pre-classification algorithms in everyday practice, very high savings on respiratory function tests can be achieved and the quality of care is not affected. In particular, savings of 98.7% of total SPIR costs (see Table 3) can be achieved on 1291 patients that were correctly classified as healthy patients. When calculating these savings, the average cost of an hourly visit to a medical specialist as well as the costs of blood gas analysis and necessary filters for SPIR tests were taken into account.

The overall EDS was tested in real-time on 1650 consecutive patients enrolled at healthcare institutions in Bosnia and Herzegovina during 2016. The performance of EDS is reported in Table 4.

Out of 1495 patients with some respiratory disease, 1442 were correctly classified resulting in sensitivity of 96.45%. Specifically, out of 859 patients with a confirmed diagnosis of asthma, 96.62% were correctly classified, while, out of the 636 COPD patients, 96.22% were correctly classified by the EDS. In addition, 98.71% of the 155 patients with normal lung function were correctly classified.

While its installment at healthcare institutions in Bosnia and Herzegovina from January to October 2016, implemented EDS results indicated high potential for use in everyday clinical practice. Within the 1495 samples with disease diagnosis, the EDS diagnosed 1123 reports without performing additional BDT/BPT tests, and when compared to diagnosis performed by medical professionals, there was a 49.23% decrease in conducting BPT/BDT tests. Furthermore, the pre-classification algorithm accuracy rate was very high, indicating that in 98.71% cases of COPD and asthma diagnosis performed by medical professionals, confirmatory pulmonary tests could have been avoided.

Based on reviewer's recommendations, after real-time validation of the system performance in healthcare institutions, several models based on Random Forests, Gradient Boosting and Logistic Regressions algorithms<sup>17–19,32–41</sup> were developed to evaluate results of Expert System relative to these machine learning algorithms. These models were developed based on 3000 samples. Machine learning models had 4 input parameters that were also used as inputs to developed neural network architecture. Weight factors of these inputs to machine learning are presented in Fig. 2.

As it can be seen from Table 5, Gradient Boosting algorithm had the best accuracy in respect to other machine learning models and developed artificial neural network. This is because Gradient Boosting finds best features and focus weight increase on misclassified data so that their importance increases and therefore entire system accuracy.

Following results from the Table 5, the future research in this area should be continued in direction of usage of machine learning techniques, and not just neural networks.

## Discussion

Due to a population lifestyle that includes smoking and air pollution, as well as other various factors, we have faced challenges for decades in diagnosis and management of various chronic respiratory diseases. Standardized procedures for respiratory disease diagnosis and management are available<sup>8,9</sup> but these procedures still have not yielded results that significantly reduce mortality rates, especially in remote areas where medical specialists are not always available. Thus, new innovative approaches are needed to help cope with this kind of disease. Considering the success of artificial intelligence and fuzzy logic had in other scientific disciplines, it was anticipated to yield similar results when applied to the problem of respiratory disease diagnosis, especially COPD and asthma. Indeed, our results show that the automated approach in disease diagnosis provides a useful solution



No. of reports		True pre-classifications	False pre-classifications	% of true pre-classifications
Disease	2735	2603	132	95.17
<b>Healthy</b>	<b>1308</b>	<b>17</b>	<b>1291</b>	<b>98.70</b>
			<b>Average:</b>	<b>96.93</b>
COPD	1028	959	69	93.33
Asthma	1707	1644	63	96.31
<b>Healthy</b>	<b>No. of samples</b>	Cost of SPIR		
	<b>1308</b>			
True pre-classifications	<b>1291</b>	98.7% of SPIR cost for 1308 patients		
False pre-classifications	<b>17</b>	1.3% of SPIR cost for 1308 patients		

**Table 3.** Pre-classification validation accuracy and cost analysis. *Cost of SPIR testing for healthy patients \$60 + the average price of a filter needed during pulmonary tests \$1 + average costs of hourly visit to medical professional \$100 roughly + additional blood gas analysis (complete panel price is about \$10 per patient) \*The costs of SPIR testing for Bosnia and Herzegovina are taken as approximate values based on various pricing from public and private healthcare institutions in June, 2017.*

	EDS output				
	Number of reports $\Sigma$ 1650	Disease (COPD & Asthma)	Healthy	Prevalence 90.61%	
True condition	Disease (COPD & Asthma) $\Sigma$ 1495 (859 asthma/636 COPD)	1442	53	True positive rate Sensitivity 96.45%	False negative rate Miss rate 3.55%
	Healthy $\Sigma$ 155	2	153	False positive rate 1.29%	True negative rate Specificity 98.71%
	Accuracy 96.66%	Positive prediction value 99.86%	False condition rate 25.73%	Positive likelihood ratio 74.76%	
		False discovery rate 0.139%	Negative predictive rate 74.27%	Negative likelihood ratio 1.34%	

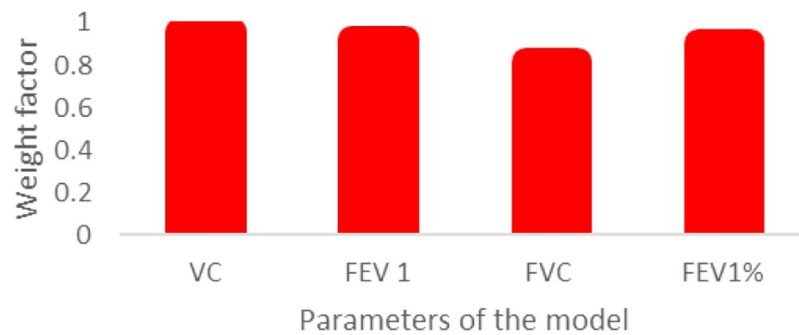
**Table 4.** System Performance.

yielding in an increase of timely correct diagnosis, especially for non-specialized medical professionals in remote areas. Based on this, early intervention programs can be developed to prevent disease complications and to ensure patients have timely delivered care.

There have been several studies focusing on diagnosing respiratory diseases using various types of artificial neural network architectures and fuzzy logic classifiers, as well as expert systems structures based on both methods. These studies have applied different neural network structures for various respiratory diseases diagnosis, including asthma and COPD, using datasets with various features<sup>42–51</sup>. Consistent with previous research in this area<sup>13–20</sup>, an automated system for COPD and asthma diagnosis based on fuzzy logic and artificial neural network was developed. Databases with previously established and confirmed diagnoses were generated for system development and patients were divided into three, mutually-exclusive categories: (1) healthy, (2) COPD, and (3) asthma. This database consisted of parameters describing patient etiology and physical examinations. During the data collection, it was noticed that patients who lived in more urban areas were more likely to have a diagnosis established earlier than those who lived in rural areas and had to go through more levels of healthcare. This means that patients from rural areas are at a higher risk of having complications and inadequate care. so Therefore, an important intervention would be to have the early diagnosis of disease at primary healthcare units which could be achieved through implementation of an automated EDS such as the one described in this research. The goal of implementing these systems in primary healthcare institutions would be to connect patients with services that would help them obtain timely care for their health and improve their access to healthcare resources.

The diversity of variables identified in the proposed EDS illustrate that medical professionals should not make a diagnosis based only on results from biological or chemical tests using laboratory values or radiological images. The patient should be evaluated as a whole system, taking into consideration functional testing factors such social factors, habits, nutrition, and epidemiological factors, as prescribed in international guidelines<sup>8,9</sup>. In this research, a total of eleven different parameters were used while performing diagnosis, with an additional five resulting from testing if needed. Patient symptoms are taken with different significance factor in the pre-classification algorithm which corresponds to international guidelines and long-term medical specialist experience. Due to the availability of spirometers and their cost, spirometry was chosen for functional testing during this research which is also in accordance with the aforementioned international guidelines, GINA and GOLD.

In assessing the strengths and weaknesses of this study, one of its major strengths is the high true positive rate of 97.58% suggesting the high EDS accuracy. As expected, artificial intelligence and fuzzy classifiers yielded high accuracy when implemented on respiratory disease. To place this result into perspective, similar performances have been achieved in other contributions<sup>32–36</sup>, but our proposed system was evaluated on both a larger dataset and real clinical settings. Furthermore, the EDS has proven reliability of results since all diagnosis used for



**Figure 2.** Weight factors of input parameters to machine learning models.

Samples Σ 3000	Model types			
	Random Forests	Gradient Boosting	Logistic Regression	Artificial Neural Network
Accuracy	97.330%	98.33%	95.33%	93.60%

**Table 5.** Comparison Of Efficiency Of Machine Learning Algorithms For Classification Of Asthma And Copd.

training were previously confirmed by trained medical specialist. This system also takes into account other patient symptoms beside the results from the functional lung test, all of which have been found to be highly associated with COPD and asthma. Finally, the real-time testing proved the utility of this system, as a simple graphical user interface system satisfied medical professional's needs.

Medical professionals included in the study had a very good response to the application. They were intrigued by the number of parameters that could be simultaneously analyzed and the time needed to generate a diagnosis. Most of medical professionals included in real-time evaluation of the system commented that it was a useful tool which helped them support the decision of patient diagnosis. Also, they all agreed that it helps reduce costs of unnecessary testing. Nevertheless, all diagnoses were finally made by the medical specialist following the recommended procedure for each patient, so false positives were treated accordingly. They undertook tests and the final diagnosis was established regardless of the system output.

As with all research, our study does have some limitations. This database consists only data collected in Bosnia and Herzegovina. It would be useful to test the system performance on data collected in different populations to evaluate the system's sensitivity. Furthermore, real time evaluation of the developed EDS was performed in only one healthcare institution in Bosnia and Herzegovina. Therefore, to further validate this automated EDS we recommend testing in other healthcare institutions.

Automated expert diagnostic systems are useful in healthcare practice because they enable diagnostics based on larger datasets and are able to take into account multiple input parameters at once. Also, the results performed by these systems are also objective and cannot be influenced by various factors. These systems are especially useful in remote healthcare institutions where a medical specialist, in this case a respiratory medical specialist, is not available and a general practitioner is not fully confident in establishing the patient diagnosis nor determining a future treatment. Accurate diagnosis of respiratory diseases in these institutions is often missed since numerous symptoms are misjudged, leading to a greater percentage of misdiagnosis or late diagnosis. These errors lead to significant negative effects on patients' health overall. The proposed EDS offers the opportunity for medical professionals to run patient symptoms through a large database of knowledge and calculate the probability of a patient having the disease. Timely diagnosis is very important step in disease treatment so in these cases, the result of EDS helps medical professional in the prescription of medication or recommendations for additional confirmatory testing. This in turn, leads to the reduction of mortality rates and the improvement of patient quality of life. It should be noted that the usage of these systems does not implicate that medical professionals are a replaceable factor in establishing diagnosis. On the contrary, these systems assist medical professionals in providing quality treatment for patients.

In the future, the authors plan to improve and validate an effective user interface for the automated EDS, enabling the usage of this system in general practitioners' offices in healthcare institutions. This system would assist general practitioners in performing preliminary diagnoses leading to optimization of time resources, decreased medical device associated costs and enhanced patient outcomes.

## Conclusion

This paper presented a novel EDS for COPD and asthma diagnosis. During development of the EDS, the system was trained with over 3000 reports acquired from 2014 to 2015 in a pulmonary clinic using the CareFusion device. Subsequently, the EDS was validated using new data acquired in a prospective study conducted at a local healthcare institution. The developed system correctly classified over 97% of the 1650 enrolled subjects, achieving a sensitivity of over 96% and specificity over of 98%.

## References

1. Tortora, G. J. & Grabowski, S. R. *Principles of Anatomy and Physiology*. (John Wiley&Sons, 2003).
2. Mayo Foundation for Medical Education and Research, Mayo Clinic: COPD and Asthma Definition. <http://www.mayoclinic.com/health/> (2017)
3. Anthonisen, N. Chronic obstructive lung disease, in *Cecil Medicine*. chapter 88 (24th ed. Goldman, L, 2007).
4. European Lung Function: COPD Burden in Europe. <http://www.european-lung-foundation.org/index.php?id=63> (2017).
5. Ferlay, J. Cancer Incidence, Mortality and Prevalence Worldwide, IARC Cancer Base. IARC Press, Lyon, France. (2004).
6. Murray, C. J. & Lopez, A. D., Alternative projections of mortality and disability by cause 1990–2020, World Health Organization, Global Burden of Disease Study. 1498–504 (1997).
7. Nielsen, K. G. & Bisgaard, H. The Effect of Inhaled Budesonide on Symptoms, Lung Function, and Cold Air and Methacholine Responsiveness in 2- to 5-year-old Asthmatic Children. *American Journal of Respiratory and Critical Care Medicine* **162**, 1500–1506 (2005).
8. Global Initiative for Chronic Obstructive Lung Disease: Pocket guide to COPD Diagnosis, Management and Prevention, [http://www.goldcopd.it/materiale/2015/GOLD\\_Pocket\\_2015.pdf](http://www.goldcopd.it/materiale/2015/GOLD_Pocket_2015.pdf) (2015).
9. Global Initiative for Asthma: Pocket guide to Asthma Diagnosis, Management and Prevention. [http://www.goldcopd.it/materiale/2015/GINA\\_Pocket\\_2015.pdf](http://www.goldcopd.it/materiale/2015/GINA_Pocket_2015.pdf) (2015).
10. Yawn, B. & Wollan, P. Knowledge and attitudes of family medical professionals coming to COPD continuing medical education. *International Journal of Chronical Obstrutive Pulmonary Disease*. **3**(2), 311–317 (2008).
11. Beverly, G. H. & Rosemary, H. An Expert Support System for Service Quality Improvement, Proceedings of the Twenty-Seventh Annual Hawaii International Conference on System Science, (1994).
12. Azaab, S., Abu Naser, S. & Sulisel, O. A proposed expert system for selecting exploratory factor analysis procedures. *Journal of the college of education*. **4**(2), 9–26 (2000).
13. Walia, N., Tiwari, S. K. & Malhotra, R. Design and Identification of Tuberculosis using Fuzzy Based Decision Support System. *Advances in Computer Science and Information Technology*. **2**, 57–62 (2015).
14. Asaithambi, M., Manoharan, S. C. & Subramanian, S. Classification of Respiratory Abnormalities Using Adaptive Neuro Fuzzy Inference System. *Int. Inf. and Database SysLecture Notes in Computer Science*. **7198**, 65–73 (2012).
15. Stavrakoudis, D., Mastorocostas, P. & Theocaris, J. A Pipelined Recurrent Fuzzy Neural Filter for the separation of Lung Sounds, IEEE International Fuzzy Systems Conference, 1–6 (2007).
16. Er, O. & Temurtas, F. A study on chronic obstructive pulmonary disease diagnosis using multilayer neural networks. *Journal of Medical Systems*. **32**(5), 429–432 (2008).
17. Mohkhtar, M. S. *et al.* Predicting the risk of exacerbation in patients with chronic obstructive pulmonary disease using home telehealth measurement data. *Artificial intelligence in medicine*. **63**(1), 51–59 (2015).
18. Prosperi, M. C., Marinho, S., Simpson, A., Custovic, A. & Buchan, I. E. Predicting phenotypes of asthma and eczema with machine learning. *BMC medical genomics*. **7**(1), S7 (2014).
19. Spathis, D. & Vlamos, P. Diagnosing asthma and chronic obstructive pulmonary disease with machine learning. *Health informatics journal*, (2017).
20. Badnjevic, A., Cifrek, M., Koruga, D. & Osmankovic, D. Neuro-fuzzy classification of asthma and chronic obstructive pulmonary disease. *BMC Medical Informatics and Decision Making Journal*. (2015).
21. Khan, J. *et al.* Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. *Nature medicine*. **7**(6), 673–679 (2001).
22. Sejdinovic, D. *et al.* Classification of prediabetes and Type 2 Diabetes using Artificial Neural Network, CMBEBIH 2017. IFMBE Proceedings. **62**, 685–689 (2017).
23. Güler, N. F., Übeyli, E. D. & Güler, I. Recurrent neural networks employing Lyapunov exponents for EEG signals classification. *Expert System s with Applications*. **29**(3), 506–514 (2005).
24. Hafezi, N. An integrated software package for model-based neuro-fuzzy classification of small airway dysfunction, ETD Collection for University of Texas, El Paso.
25. Zolnoori, M., Zarandi, M., Moin, M. & Teimoran, S. Fuzzy Rule-Based Expert System for Assessment Severity of Asthma. *Journal of Medical Systems*. **36**, 1707–1717 (2010).
26. Mishra, N., Singh, D. B. V., Bandil, M. K. & Sharma, P. Decision Support System for Asthma. *International Journal of Information and Computation Technology*. **3**, 549–554 (2013).
27. Uncu, U., Koklukaya, E. & Gencsoy, A. A fuzzy rule-base model for classification of spirometric FVC graphs in chronic obstructive pulmonary diseases, 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Buenos Aires, Argentina. 3866 – 3869 (2010).
28. Anand, S. K., Kalpana, R. & Vijayalakshmi, S. Design and Implementation of a Fuzzy Expert System for Detecting and Estimating the Level of Asthma and Chronic Obstructive Pulmonary Disease, Middle-East. *Journal of Scientific Research*. **14**, 1435–1444 (2013).
29. Sheikhtaheri, A. & Sadoughi, F., Developing and Using Expert Systems and Neural Networks in Medicine: A Review on Benefits and Challenges, *Journal of Medical Systems*. (2014).
30. Byvatov, E., Fechner, U., Sadowski, J. & Schneider, G. Comparison of support vector machine and artificial neural network System s for drug/nondrug classification. *Journal of Chemical Informatics and Computing*. **43**(6), 1882–1889 (2003).
31. Gnana Sheela, K. & Deepa, S. N. Review on Methods to Fix Number of Hidden Neurons in Neural Networks, Mathematical Problems in Engineering (2013).
32. Farlow, E. C. *et al.* A multi-analyte serum test for the detection of non-small cell lung cancer. *British journal of cancer* **103**(8), 1221 (2010).
33. Bureau, A. *et al.* Identifying SNPs predictive of phenotype using random forests. *Genetic epidemiology* **28**(2), 171–182 (2005).
34. Smolinska, A. *et al.* Profiling of volatile organic compounds in exhaled breath as a strategy to find early predictive signatures of asthma in children. *PloS one* **9**(4), e95668 (2014).
35. Phillips, C. O. *et al.* Machine learning methods on exhaled volatile organic compounds for distinguishing COPD patients from healthy controls. *Journal of breath research* **6**(3), 036003 (2012).
36. Panganiban, R. P. *et al.* Circulating microRNAs as biomarkers in patients with allergic rhinitis and asthma. *Journal of Allergy and Clinical Immunology* **137**(5), 1423–1432 (2016).
37. Amaral, J. L., Lopes, A. J., Faria, A. C. & Melo, P. L. Machine learning algorithms and forced oscillation measurements to categorise the airway obstruction severity in chronic obstructive pulmonary disease. *Computer methods and programs in biomedicine* **118**(2), 186–197 (2015).
38. Edmiston, J. S. *et al.* Gene expression profiling of peripheral blood leukocytes identifies potential novel biomarkers of chronic obstructive pulmonary disease in current and former smokers. *Biomarkers* **15**(8), 715–730 (2010).
39. Swaminathan, S. *et al.* Smart Asthma: A Triage Application That Leverages Machine Learning Predictions. In B74. Pulmonary Rehabilitation For Non-Copd Diagnoses (pp. A4287-A4287). American Thoracic Society (2017).
40. Soriano, J. B., Zielinski, J. & Price, D. Screening for and early detection of chronic obstructive pulmonary disease. *The Lancet*, 374(9691), 721–732 (2009).
41. Zarandi, F., Zolnoori, M., Moin, M. & Heidarnajad, H. A Fuzzy Rule-Based Expert System for Diagnosis of Asthma. *Transaction E: Industrial Engineering* **17**, 129–142 (2010).



42. Coppini, G., Miniati, M., Paterni, M., Monti, S. & Ferdeghini, E. M. Computer aided diagnosis of emphysema in COPD patients: Neural network based analysis of lung shape in digital chest radiographs, *Medical Engineering and Physics*, 76–86 (2007).
43. Price, D. B. *et al.* Symptom-based questionnaire for identifying COPD in smokers. *Respiratio*. **73**, 285–294 (2006).
44. Yawn, B. P., Mapel, D. W. & Mannino, D. M. Development of the Lung Function Questionnaire (LFQ) to identify airflow obstruction. *International Journal of Chronical Obstructive Pulmonary Disease*. **5**, 1–10 (2010).
45. Hanania, N. A., Mannino, D. M. & Yawn, B. P. Predicting risk of airflow obstruction in primary care: validation of the lung function questionnaire (LFQ). *Respiratory Medicine*. **104**, 1160–1170 (2010).
46. Martinez, F. J., Raczek, A. E. & Seifer, F. D. Development and initial validation of a self-scored COPD Population Screener Questionnaire (COPD-PS). *COPD*. **5**, 85–95 (2008).
47. Belos, C. C., Papadopoulos, A., Rosso, R. & Fotiadis, D. I. Identifying COPD Patients' Health Status Using an Intelligent System in the CHRONIOUS Wearable Platform. *IEEE Journal of Biomedical and Health Informatics*. **18**, 731–738 (2013).
48. Badnjevic, A., Gurbeta, L., Cifrek, M. & Marjanovic, D. Classification of Asthma Using Artificial Neural Network, IEEE 39th International convention on information and communication technology, electronics and microelectronics (MIPRO) (2016).
49. Badnjevic, A., Gurbeta, L., Cifrek, M. & Marjanovic, D. Diagnostic of Asthma Using Fuzzy Rules Implemented in Accordance with International Guidelines and Physicians Experience, 39th International Convention on Information and Communication Technology, Electronics and Microelectronics (IEEE MIPRO) (2016).
50. Granulo, E., Becar, L., Gurbeta, L. & Badnjevic, A. Telemetry System for Diagnosis of Asthma and Chronical Obstructive Pulmonary Disease (COPD), Internet of Things Technologies for HealthCare. Mobyen Uddid Ahmed, Shahina Begum, Wasim Raad. Springer International Publishing, pp. 113–118, [https://doi.org/10.1007/978-3-319-51234-1\\_18](https://doi.org/10.1007/978-3-319-51234-1_18) (2016).
51. Gurbeta, L., Badnjevic, A., Maksimovic, M., Omanovic-Miklicanin, E. & Sejdic, E. A telehealth system for automated diagnosis of asthma and chronical obstructive pulmonary disease. *Journal of the American Medical Informatics Association* (2018).

## Acknowledgements

The authors would like to thank the Respiratory Medicine Specialists from the Hospital Sarajevo, Bosnia and Herzegovina for their help during the research and CareFusion Company for providing data for the system development.

## Author Contributions

A.B. Coordinated the study, designed the experimental set-up, build the Expert System, did validation and testing of System and assisted in writing manuscript. L.G. Did data analysis and wrote the manuscript. E.C. Contributed to the manuscript and provided critical insights. All authors read and approved the final manuscript.

## Additional Information

**Competing Interests:** The authors declare no competing interests.

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



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018