

## ORIGINAL PAPER

# Developing an Intelligent System for Diagnosis of Asthma Based on Artificial Neural Network

Behrouz Alizadeh<sup>1</sup>, Reza Safdari<sup>2</sup>, Maryam Zolnoori<sup>3</sup>, Azadeh Bashiri<sup>4</sup><sup>1</sup>School of Allied-Medical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran<sup>2</sup>Health Information Management Department, School of Allied-Medical Sciences, Tehran University of Medical Sciences, Tehran, Iran<sup>3</sup>School of Informatics and Computing, Indiana University, Indiana, USA<sup>4</sup>School of Allied Medical Sciences, Tehran University of Medical Sciences, Tehran, Iran

Corresponding author: Azadeh Bashiri. School of Allied-Medical Sciences, Tehran University of Medical Sciences, Tehran, Iran, E-mail: bashiri.azadeh@gmail.com

**ABSTRACT**

**Introduction:** Lack of proper diagnosis and inadequate treatment of asthma, leads to physical and financial complications. This study aimed to use data mining techniques and creating a neural network intelligent system for diagnosis of asthma. **Methods:** The study population is the patients who had visited one of the Lung Clinics in Tehran. Data were analyzed using the SPSS statistical tool and the chi-square Pearson's coefficient was the basis of decision making for data ranking. The considered neural network is trained using back propagation learning technique. **Results:** According to the analysis performed by means of SPSS to select the top factors, 13 effective factors were selected, in different performances, data was mixed in various forms, so the different modes was made for training the data and testing networks and in all different modes, the network was able to predict correctly 100% of all cases. **Conclusion:** Using data mining methods before the design structure of system, aimed to reduce the data dimension and the optimum choice of the data, will lead to a more accurate system. So considering the data mining approaches due to the nature of medical data is necessary.

**Key words:** Asthma, Data mining, Artificial Neural Network.**1. INTRODUCTION**

Asthma is one of the most common diseases throughout the world. And in different parts of the world, has a prevalence of 1.4 to 27.1 percent (1, 2). Asthma is a chronic reversible airway disease and usually gives a good response to treatment (3). Diagnosis of this disease is based on GINA criteria and clinical findings, including coughing, shortness of breath, heaviness and wheezing occurs (4). Asthma has many similarities with other respiratory diseases (5), and also has relationships with other diseases such as systemic amyloidosis and poly chondroitin recurrent (6). Some of the obstructive diseases of the upper respiratory tract have symptoms such as vocal cord dysfunction similar to asthma (7). Many physicians understand wheezing and shortness of breath and along with, cough equivalent to diagnose asthma. These symptoms are not specific to asthma and wrong diagnosis of it and inappropriate treatment can lead to physical and financial complications (5). The rapid development of computer technology has led to the automation of many processes that were previously done by professionals. Technique development based on artificial intelligence (AI) as a major is an important requirement to solve complex problems. Medical diagnosis, due to the complexity of the human body and mind and vague knowledge in this field, is a good example of a class of complex problems and also the existent knowledge are different between people with various experiences. The physicians' experience role in the diagnostic process is undeniable, and to avoid wasting time in the diagnosis, techniques based on artificial intelligence are required, so that detect it so rapid, reli-

able, accurate and knowledge base (8). In particular, artificial intelligence is a branch of data mining that uses automate process for finding information and patterns between the data. Data mining is used in a wide range of various purposes, but one of its most obvious application, is in the process of medical diagnosis based on clinical findings, which is used in order to explore the relationship between clinical findings and diagnostic results. Designing optimized intelligent systems in addition to the structure of the system depends on the input data. The garbage in garbage out phrase particularly emphasized this point that achieving to optimal system, accurate information and best use of the relevant data, are necessary (8, 9). Data mining issues often involve the kind of points that they all can be as a potential of predictor of outputs, but in practice, a few of them in different levels influence the output. In order to reduce input dimensions and selection of optimum alternatives, there are several ways, but the common stages among these methods are as follows:

- \* Delete the unimportant data,
- \* Ranking the remained data,
- \* Select among the ranked data (9).

One of the artificial intelligence techniques to solve real-world problems is, using the concept of a neural network. The main idea to use a neural network is, learning ability by old data and derives a new solution for overcoming the problems (10). The purpose of this study is performing the data mining on database of clinical results for excellent and effective selections in the diagnosis of asthma and making intelligent systems using the neural network technique, that is very useful

Symbol	Concept	explanation
cough	cough	Based on the type, frequency and intensity is placed in one of the 0 to 5 groups.
phlegm	Phlegm	Is classified according to the presence or absence of sputum
Wheeze	Wheeze	Wheezing while breathing (0-3)
DYSP	Respiratory distress	Based on type and frequency of wheezing is placed in one of the 0 to 3 groups.
CT	pursiness	
Exe.s	Symptoms of exercise-induced	
AR	Nasal mucosal swelling caused by allergies.	
BMI	Body Mass Index	Is used To distinguish overweight patients from other patients.
GER	Gastric esophagitis reflux	
Par2	Allergies /Catarrh of Allergy in both parents	
Par1	Allergies /Catarrh of Allergy in one of the parents	
Rel	Allergies /Catarrh of Allergy in one of the relatives	
R.AL	Response to allergens	
R.Ir	Response to irritants	
R.Em	Emotional reactions	
Ecz	Background of eczema or Catarrh	
Hosp	Hospitalization Before the 3 years old due to viral infections	
Food	Food allergy before the 3 years old	
AP	air pollution of the Location	
P.smoke	smoking Now or in the past	
damp	Humidity levels in the home or humid environment	
cold	Colds	

Table 1. Clinical characteristics used by physicians for asthma 'diagnosis

to reduce the financial and repeated examination costs, and also the wrong diagnosis and its consequences.

## 2. METHODOLOGY

Investigating community in this Study were the patients who had visited in one of the lung clinics in Tehran that due to the completeness of the recorded information in the case records, selected 254 patients. This information in a database can be found as clinical findings related to the asthma and also the diagnosis results in either the presence or absence of asthma, were stored. The data were analyzed using SPSS statistical software. Basis of decision making for ranking data was the Pearson chi-square coefficient, which aims to identify the effective characteristics on decision making of the diagnosis. Table 1 shows Clinical characteristics used by physicians to diagnose asthma.

Given the importance of patient characteristics as input to optimal neural network, the data in the database were investigated in terms of optimality and effectiveness of the output of

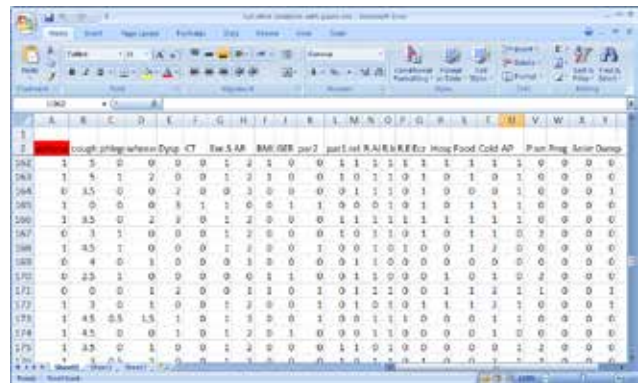


Figure 1. used clinical Results in detecting asthma

a statistical method and based on the Pearson chi-square coefficients. Due to the fragmenting the data before inserting into the database and under the conditions outlined in Figure1, using the chi-square distribution is as reasonably practical criterion for ranking data.

The basis of this test, is a comparison between the expected and calculated value. It means, we want to know that is there any significant difference between the observed and expected frequencies or the difference is negligible and the result of chance. In fact, we want to know that is there a relationship between two variables or they are independent. In the below formulas, Fo is observed frequency and Fe is expected frequency that calculated as follows in Equation 1. Also feature selection method based on criteria introduced in relationship that calculated as follows in Equation 2.

$$Fe_i = \frac{(n_i \times n_j)}{n}$$

Equation 1. Expected frequency

$$\chi^2 = \sum_{i=1}^k \frac{(Fo_i - Fe_i)^2}{Fe_i}$$

Equation 2. Chi square value computation for observations and expected values

According to table 2, unimportant data are excluded from total results and the first 13 cases were selected as the factors that affect asthma detecting, and the used database was reduced to 13 items for each patient. Randomly, 70% of the data are classified as training data and the rest as testing data of the proposed system. In designing the neural network, the programming language C # fourth version of Visual Studio 2010 and the neural network Library of 4 NeuronDotNet that is free and is made in implementing the neural network in dotNET, has been used. Such as figure 2, the network layers were implemented as input, hidden and output layer.

Input layer is linear and is only responsible for data transferring from the internal structure of the network; hidden and outlet layers are Sigmoid type, and perform calculations on the network. Network structure was designed as13-26-1 in which there are 13 nodes in the input layer, 26 nodes in the hidden layer and one node in the output layer. The output classification criterion is as a conditional function according to Equation 3. In this Equation, 1 means affected by the disease and 0 means the absence of the disease. The considered

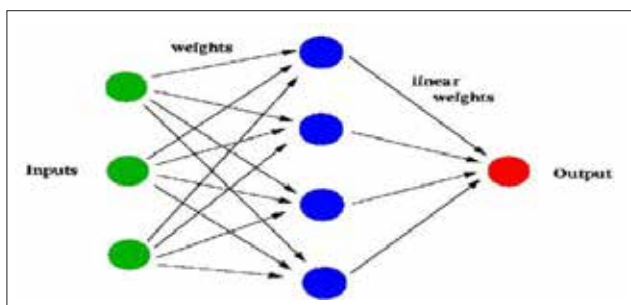


Figure 2: An example of neural network with input, hidden and output layers

neural network is trained using back propagation learning technique.

$$Y = \begin{cases} 1 & \text{output} > 0.5 \\ 0 & \text{output} \leq 0.5 \end{cases}$$

Equation 3. Converting the network's outputs in binary mode

### 3. RESULTS

From the total of 254 recorded cases related to research, 169 patients suffering from asthma and 85 cases were not affected with this disease. In the records related to the asthmatic patients, 22 factors are recorded in relation to the diagnosis. According to the performed analysis by means of SPSS to select the top factors, 13 factors were effective.

Designed neural networks with a different numbers Epoch were trained and evaluated. Obtained results for different number of EPOCH are as Table 3. Eventually, after training to 100000 EPOCH, the mean square error was reduced to less than 0.00002. Given that the data for training and testing the neural network are selected by random Shuffleboard (mixing and shuffle) in different runs. Due to the nature of Shuffleboard, different modes of network training and testing data,

Rank	Field	Importance	Value
1	cough	important	1.0
2	Exe S	important	1.0
3	Damp	important	1.0
4	R Em	important	1.0
5	AR	important	1.0
6	wheeze	important	1.0
7	Dysp	important	1.0
8	Hosp	important	1.0
9	R Ir	important	1.0
10	R AL	important	1.0
11	phlegm	important	1.0
12	par2	important	0.998
13	CT	important	0.961
14	rel	Marginal	0.947
15	Ecz	Marginal	0.944
16	par1	Unimportant	0.881
17	AP	Unimportant	0.867
18	P smoke	Unimportant	0.792
19	GER	Unimportant	0.76
20	Feed	Unimportant	0.768
21	BMI	Unimportant	0.645
22	Cold	Unimportant	0.083

Table 2. Ranking and selection of the top clinical findings in asthma detecting

created, and in all cases system was able to predict 100% correct.

Minimum and maximum values for networks accuracy of different performances and the results of Shuffleboard func-

The most accuracy	Minimum accuracy	Number of Epoch
100%	93.5%	1000
100%	96.5%	10000
100%	97.4%	20000
100%	98.6%	100000

Table 3. EPOCH number and accuracy of the designed neural network

tion impact, have been reported.

In the higher number of EPOCH, the results for the lowest and highest levels of decision accuracy remains constant and it didn't need more training to system and a further number of EPOCHS. In the below chart, the vertical axis represents the mean square error and the horizontal axis indicates Education

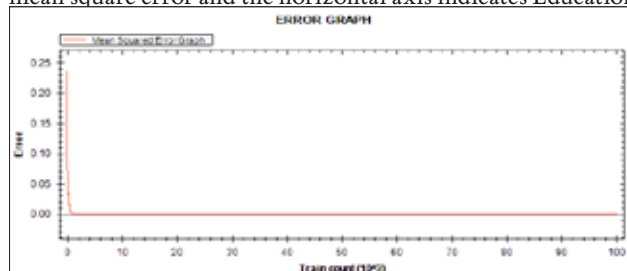


Figure 3. The neural network error rate

Epochs' of Network in 100000 EPOCH.

### 4. DISCUSSION AND CONCLUSION

The aim of this study is to develop an intelligent system for detecting the presence or absence of asthma in which the clinical features in terms of influencing the output are investigated and ordered, and only the effective features are used as the input of neural network. 70% of the data have been trained as the required networks' input to achieve the desired result and reduce errors, and the rest of data are used to test the system. As a result, the system is able to accurately predict disease based on clinical findings in 100% of the cases. To study former neural networks in the field of medical science, In a study that was conducted in 2012 by Zonnour and colleagues entitled computerized intelligent system for detecting asthma in children, the fuzzy intelligent system designed to predict Asthma and was able to predict 100% correctly sick and healthy cases (2). Biglarian and colleagues' research which compared the ANN model and Cox regression on predicting survival of patients with gastric cancer can be pointed, it was done in 2011 and was achieved to 81.511% accuracy rate (12). In a research that Sedhi and colleagues were done in 2010, called designing integrated artificial neural network to predict the metabolic syndrome and insulin resistance index, they used neural networks for prediction of the syndrome and reached to 75.67% accuracy rate (13). Research that was conducted by Zanganeh, called the IHD analysis by data mining methods, the proposed system was able to diagnose the patients and healthy cases 96.4% accurately (14). In a study of Alfred Bvshlvyr and colleagues as intelligent neural networks that to predict lymph node metastasis in gastric cancer conducted in 2004, the proposed system was able to predict accurately in 93% cases (15). Since the one of the most important factors in the development of the neural network is to select the appropriate structure for it and using optimal data input, using the waste data will lead to incorrect education of neural network and will put it in the wrong path. Also by select the optimum data as input of the network by statistical methods, the number of data that has most efficiency in outlets be selected. This action will leads to a proper education to the network and will increase the accuracy of its decision making. Finally, the neural network was able to correctly identify in 96.5% to 100% of cases. Achieving to mentioned accuracy rate, depends on different factors, including the nature of the

disease and its risk factors and the clinical features, but one of the most important findings in achieving to the accuracy is to measure the optimality of the effects of finding on the obtained results of detection and ordering and selection of priority cases. According to the obtained results in the selection of effective characteristics, among the 22 cases recorded in the patients' records, 13 agents were selected as the most effective factors for detecting asthma. It seems that practitioners' focus on selected cases to diagnosis the asthma, increases accuracy and certainty in decision making. The success of any intelligent system, regardless of its internal structure and the used techniques for Intelligent, is largely depends on the input data optimality. Therefore, commonly, using data mining methods before designing the structure of system, aimed at reducing the data dimension and optimal choice, will lead to a more precise system. So, using data mining approaches is essential due to the nature of medical data. According to low error rate of the proposed system in detecting asthma based on clinical findings, above system can be used as an assistant of decisions on detecting disease.

**CONFLICT OF INTEREST: NONE DECLARED.**

## REFERENCES

1. Fauci A, Braunwald E, Kasper D, Hauser S, Longo D, Jameson J, et al. *Harrison's Principles of Internal Medicine*, 17th Edition: McGraw-Hill Companies, Incorporated, 2008.
2. Zolnoori M, Fazel Zarandi MH, Moin M, Heidarneshad H, Kazemnejad A. Computer-aided intelligent system for diagnosing pediatric asthma. *Journal of medical systems*. 2012; 36(2): 809-22. <http://dx.doi.org/10.1007/s10916-010-9545-5>
3. Murray JF, Mason RJ. *Murray and Nadel's textbook of respiratory medicine*. Philadelphia, PA: Saunders/Elsevier, 2010.
4. Mehrabi SSM, Ghauomi SMA. Determining the Most Suitable Spirometric Parameters to Differentiate Chronic Obstructive pulmonary Disease (COPD) from Asthma. *Armaghan Danesh*. 2010; 14(4): 76-86.
5. Tilles SA. Differential diagnosis of adult asthma. *The Medical clinics of North America*. 2006; 90(1): 61-76. <http://dx.doi.org/10.1016/j.jacr.2009.03.005>
6. Sharma SK, Ahluwalia G, Ahluwalia A, Mukhopadhyay S. Tracheobronchial amyloidosis masquerading as bronchial asthma. *The Indian journal of chest diseases & allied sciences*. 2004; 46(2): 117-119. <http://dx.doi.org/10.1097/LBR.0b013e318166d233>
7. Suri JC, Sen MK, Chakrabarti S, Mehta C. Vocal cord dysfunction presenting as refractory asthma. *The Indian journal of chest diseases & allied sciences*. 2002; 44(1): 49-52.
8. Innocent PR, John RI, Garibaldi JM. Fuzzy methods for medical diagnosis. *Applied Artificial Intelligence*. 2004; 19(1): 69-98.
9. Agrawal R, Gehrke J, Gunopulos D, Raghavan P. *Automatic subspace clustering of high dimensional data for data mining applications*: ACM, 1998.
10. Grochowski M, Duch W. *Constructive neural network algorithms that solve highly non-separable problems*. Constructive neural networks: Springer, 2009: 49-70.
11. Azar A, Momeni M. *Statistics and its Application on Management*. Tehran: SAMT, 2012.
12. Biglarian A, Hajizadeh E, Kazemnejad A. Comparison of artificial neural network and Cox regression models in survival prediction of gastric cancer patients. *Koomesh*. 2010: 215-220.
13. MSO et al. Designing artificial neural networks for predicting the metabolic syndrome associated with insulin resistance index: Tehran Lipid and Glucose Study. *Daneshvar*. 2009; 17 (85): 29 -38.
14. Zanganeh S. *Data mining techniques to analyze cardiac ischemia*: Science and Research Branch, Tehran, 2012.
15. Bollschweiler EH, Monig SP, Hensler K, Baldus SE, Maruyama K, Holscher AH. Artificial neural network for prediction of lymph node metastases in gastric cancer: a phase II diagnostic study. *Annals of surgical oncology*. 2004; 11(5): 506-511. <http://dx.doi.org/10.1245/aso.2004.04.018>