

## **Prognostic system for early diagnosis of pediatric lung disease using artificial intelligence.**

**Juliet Rani Rajan<sup>1</sup>, Chilambu Chelvan A<sup>2</sup>**

<sup>1</sup>Research Scholar, Satyabhama University, Infosys Technologies, India.

<sup>2</sup>Professor and Head, Department of Electronics and Instrumentation, RMD Engineering College, Chennai, India.

### **Abstract**

**With the huge growth in the volume of data today, there is an enhanced need to extract meaningful information from the data. Data mining contributes towards this and finds its application across various diverse domains such as in information technology, retail, stock markets, banking and healthcare among others. The surge in population together with the growth in diseases has obliged the insertion of data mining in medical diagnosis to pull out the underlying hidden pattern. Of these, asthma is a disease of high prevalence among children. Asthma prevalence increased from 2001 to 2010 and is now at its highest level. Asthma prevalence was higher among females, children and those with family income below the poverty level. It may be difficult to tell whether your child's symptoms are caused by asthma or something else. Asthma, a disease which is extremely reliant on historical data for early diagnosis, has inclined researchers to follow the data mining techniques for the pre-diagnosis procedure. Existing medical techniques like X-Ray, Spirometer, impulse oscillometry and other lung examination techniques not only require complex equipment and high cost but are also proven to be efficient only when the child is more than 5 years old. The proposed system involves in developing a data mining model that will aid in the grouping of patients into the set that could hypothetically test positive for asthma. Depending on the results obtained as part of pre-diagnosis process from the tool, the doctor can perform the diagnosis for the confirmation of asthma in the patient and initiate the treatment at an early stage.**

**Keywords:** Artificial neural networks, Biomarkers, Data mining, Pattern evaluation.

*Accepted November 30, 2016*

### **Introduction**

An asthma flare-up, which some people call an asthma attack or episode, happens when a person's airways get swollen and narrower and it becomes a lot harder for air to get in and out of the lungs. Asthma affects about 1 or 2 kids out of 10. Asthma is one disease that children start to develop even before the age of 5. This asthma is very difficult to diagnose during the young age. Most of the time it is difficult to distinguish between asthma and other childhood condition since the symptoms of either of the conditions can be similar [1]. Not all children have the same kind of asthma symptoms, these symptoms can vary from episode to episode in the same child. Children aged 0–17 years had higher asthma prevalence (9.5%) than adults aged 18 and over (7.7%) for the period 2008–2010. Females had higher asthma prevalence than males (9.2% compared with 7.0%) [2].

Asthma can be difficulty diagnosed, especially in young

children, who may have wheezing and coughing as part of a viral illness, such as bronchiolitis. If the child is old enough, he/she may have the ability to take up any test to diagnose asthma. For children around 5 year of age can undergo the spirometer test to measure lung function. Another test that children around 5 years of age can perform is to measure the exhaled nitric oxide which is a marker for airway inflammation. Impulse oscillometry which is used to measure airway resistance is used in children who cannot perform the other lung tests [3]. These are some of the tests that are commonly used to diagnose lung disease and are rarely ordered in the diagnoses. The enormous number of younger children is diagnosed based solely on the historical data. Asthma is an exceptionally manageable illness. Definitely, for most children suffering from lung disease, control is so active that it extents to a virtual cure [4]. Earlier diagnosis plays a major role for the virtual cure.

As the dimensions of data are mounting proportionally

with the increase in population, there is a greater need to extract the knowledge from the data. Data mining contributes much towards this and finds its application in various diverse fields including the healthcare industry. Data mining is the process of sifting through historical data thus providing an insight into the patterns from large dataset and helps to incorporate the pattern in everyday activity. Data mining aids in medical diagnosis to figure out the underlying pattern for the cause of the disease. Researchers are suggesting that applying data mining techniques in identifying effective pre-diagnosis of the disease can improve practitioner performance [5,6]. Lung cancer being a disease which is highly dependent on historical data can make use of data mining for its early detection. Researchers have been investigating on applying various data mining techniques on lung cancer dataset for early diagnosis of lung cancer.

This paper proposes a model for measuring if applying data mining techniques to pediatric asthma dataset can provide reliable performance in the early detection of asthma. The rest of the paper is structured as follows: Section 2 provides a literature survey on using data mining techniques which help health care professionals in the early diagnosis of asthma in children below five years of age.

### **Literature Survey**

Over the last decade, many remarkable procedures of data mining were offered to detect various types of lung diseases. A small number of the methods are described below with their impact and margins.

Chatzimichail et al. [7] proposed a method for the prediction of persistent asthma in children. These method pacts with training a Multi-Layer Perceptron network with the back propagation network as the learning algorithm. This method recognizes three classes of asthma severity through the results of breathing tests. Here, 200 children-patient cases were taken into account and the age of the children were between 10 and 12. Depending on the severity, mild, moderate and severe, the appropriate course of medical treatment is given to the patient. This technique can be applied on patients who is already suffering from asthma, rather than diagnosing the disease at an early stage thus controlling the symptom. The learning method employed here is back propagation algorithm which is a supervised learning method which tries to perform the classification based on the input and target values with which it is trained. Any occurrence of new symptoms cannot be adapted in this type of supervised network.

Ansar et al. [8] proposed a method for automatic diagnosis of asthma using neurofuzzy system. The architecture proposed by them is capable of dealing with complex system which includes rule chaining. The architecture incorporates the adaptability of back propagation bases neural network and the decision making capability of a rule based fuzzy system. Here, the factors considered for the diagnosis are age, gender, economic status and alcohol and smoke intake. Hence, the diagnosis is restricted only

to these four factors which do not always give a proper accuracy.

Saumya et al. [9] proposed a method for asthma detection by breathe analysis. This is done with the help of an electronic nose which is made of an array of sensors. The sensor measures the level of Nitric Oxide present in the exhaled breathes which is one major biomarker of asthma. The breathe of the patient is captured in a gas bag and then subjected to the sensor. Though NO is responsible for airway inflammation, Published evidence was inconclusive as to whether Fen (Fractional Exhaled NO) is a useful management strategy for asthma [10]. It is proven that both Fen and induced sputum are useful asthma biomarkers. Simply measuring the level of Fen will not yield the desired result.

Artificial Neural Networks (ANN) provide a precise and a prevailing tool that help doctors to evaluate, model and finally figure out the sense of multifaceted clinical facts across an extensive array of medical applications [11]. It is basically a mathematical model developed on the basis of biological neural networks. Each node in the input layer represents each attribute of the patient's dataset. The values from the input layer are then transmitted to the nodes in the hidden layer along with the weight values where the learning takes place. Upon the completion of the learning process, the classification is performed in the output layer.

Economou et al. [12] has come up with a model that diagnosis pulmonary disease such as tuberculosis, asthma, lung cancer, occupational disorders of lungs and others. They also proved that the feed forward networks, especially the back propagation network and the Kalman filter give a better performance.

Several neural networks approaches were compared by Juliet et al for the pre-diagnosis of lung cancer [13]. It has been concluded that Self Organizing Map could give better classification accuracy. This is then proved by Juliet et al. [14] with the help of Gene Expression Dataset.

### **Materials and Methods**

#### **Data Source**

The dataset of children population sample during enrolment at the hospital can be analyzed. The history of the patient, along with other prognostic factors is captured from the patients by means of questionnaires. A total of 70 records have been gathered. These records were a combination of both asthma and non-asthmatic patients. Out of these 70 records, 40 records were used for training purpose and the remaining 30 records were kept as testing set. Some of the prognostic factors included in the questionnaire are:

- Age
- Sex
- Pregnancy duration
- Birth weight
- Family history of asthma

- Economic status
- Second hand smoke
- Smoking of mother during frequency
- Eczema
- Food allergy

Many prognostic factors were identified and finally the below factors were shortlisted for the training process in Table 1.

## Experimental Evaluation

### Back Propagation Network

The Back Propagation (BP) neural network algorithm is a feed forward network with a multi-layered organization which is trained and it is one of the most extensively practical neural network models. The network learns based on the training records by propagating the errors. The pattern learnt during the training can be used to classify the test data. The network thus can be used to learn and store a huge amount of input-output mapping relations. The disclosure of mathematical equation used to disclose the mapping relations is not required in advance. Its learning rule adopts the steepest descent method in which the back propagation regulates the weight value and threshold value of the network so that the minimum error sum of square is achieved [15-17].

In the network, each node represents each attribute. The patient information is provided into the network in a way that each feature figures out its position into the analogous node. The nodes in the hidden layer and the learning adjust itself to come up with a good mapping relation between the given input and the output. The learning ceases once it reached the saturation point. Care is taken to see that the learning is not stopped at the local minima. The error value is calculated and it is transmitted back to the network and the weight values are adjusted so that the overall error can be minimized.

$$MSE = 1 / Nn = 1N(d(n) - y(n))^2$$

**Table 1.** Prognostic factors for the training process

S. No.	Category	Prognostic Factor
1	Demographic Details	Age, Sex, Height, Weight
2	Neonatal Characteristics	Birth weight, Pregnancy duration, Mother smoking during pregnancy, Breast fed, Exposure to second hand smoke during pregnancy
3	Bronchiolitis Episodes	Frequency of Respiratory Infection, Frequency of cold and cough, Frequency of Wheeze
4	Allergy Symptoms	Food allergy, Eczema, Pharmaceutical allergy
5	Family History	Family history of asthma, Presence of asthma in mother, Presence of asthma in father.

Where N represents the number of records, d (n) is the expected output and y (n) symbolizes the target output.

### Algorithm

The algorithm for pediatric asthma works as follows [16]:

1. The raw patient record obtained has been cleaned.
2. Encode the attributes of the record with the necessary values for every prognostic factor. The values are then normalized.
3. Apply the encoded record as input to the network and work out the output. Initial weights are random numbers.
4. Calculate error for neuron B.

$$Error = Output (1 - Output B) (Target - Output)$$

5. Update the weight value. Let W+AB be the new (trained) weight and WAB be the initial weight.

$$W+AB=WAB+(Error \times Output A)$$

6. Calculate the Errors for the hidden layer neurons.

$$Error A = OutputA (1 - Output A)$$

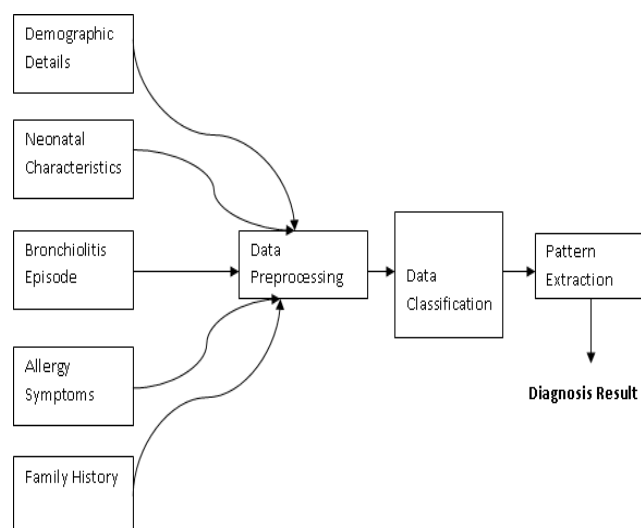
$$(Error WAB + Error C WAC)$$

By repeating this method we can train a network of any number of layers. The network is trained for global minima.

Once the training is completed, the network can be analyzed to extract the pattern which will be useful for the medical practitioner to analyses the result for further treatment. The architecture of the process is illustrated in the Figure 1.

## Results and Discussion

The implementation of back propagation network in pediatric asthma dataset has been implemented on Matlab. The pediatric dataset consists of 70 patient records. Each of the 70 patient records contains 12 attributes. Mean is used to fill the missing data. The number of training data



**Figure 1.** Prognostic model for pediatric asthma diagnosis

**Table 2. Implementation results**

	Patient Records	Classification Accuracy
Training Data	40	-
Testing Data	30	100%

taken was 40 records and the remaining 30 records were used for the testing process. The classification accuracy was found to be 100%. The bias term is taken as -1.0740. Table 2 shows the implementation result.

For the records given as input to the back propagation network, it has been observed that the greatest weight value is associated with the family history of asthma followed by the exposure to second hand smoke during pregnancy. The least weight value was noticed in the height of the patient. The model performs a binary classification: positive pediatric asthma and momentary allergy. After training the 40 records, it was then tested with the remaining 30 records. The classification accuracy for the test record was 100%. This model can be used as a pre-diagnoses tool for the practitioners.

### Conclusion

Because asthma is a disease which is greatly dependent of pre-historic data, there is a highly probability of diagnosing pediatric asthma based on experience which accounts to the number of records giving as training samples to the network. This work has demonstrated a technique of pre-diagnosing the disease based on back propagation learning algorithm where in the weight values of the nodes get adjusted till the desired output has been obtained. Utmost care has been taken to ensure that the training is done till a global minima is reached. This model can help the doctors in the pre-diagnosis process of childhood asthma thereby preventing the disease during the early itself.

### References

1. [http://kidshealth.org/kid/asthma\\_basics/what/asthma.html](http://kidshealth.org/kid/asthma_basics/what/asthma.html)
2. <http://www.who.int/mediacentre/factsheets/fs206/en/>
3. [http://www.medicinenet.com/asthma\\_in\\_children](http://www.medicinenet.com/asthma_in_children)
4. <http://www.scientificamerican.com/article.cfm?id=can-asthma-be-cured-what&page=2>
5. Brian R, Nancy W. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: Methods of a decision-maker-research

partnership systematic review. *Implementation Science* 2010; 5: 12.

6. Qeethara A. Artificial neural networks in medical diagnosis. *International Journal of Computer Science Issues* 2011; 8: 1-2.
7. Chatzimichail EA, Rigas AG, Paraskakis EN. An artificial intelligence technique for the prediction of persistent asthma in children. *Information Technology and Applications in Biomedicine (ITAB). 10th IEEE International Conference on* 2010; 1: 3-5.
8. Ansari AQ, Gupta NK, Ekata E. Automatic diagnosis of asthma using neurofuzzy system. *Computational Intelligence and Communication Networks (CICN). Fourth International Conference* 2012; 819: 3-5.
9. Shelat SJ, Patel HK, Desai MD. Breathe analysis by electronic nose for asthma detection. *Engineering (NUiCONE), 2012 Nirma University International Conference* 2012; 1: 6-8.
10. Leung TF, Ko FW, Wong GW. Recent advances in asthma biomarker research. *Ther Adv Respir Dis* 2013; 7: 297-308.
11. Qeethara A. Artificial neural networks in medical diagnosis. *International Journal of Computer Science Issues* 2011; 8: 1-2.
12. Economou GPK, Spiropoulos C, Economouopoulos NM, et al. Medical diagnoses and artificial neural networks: A medical expert system applied to pulmonary diseases. *IEEE Conference* 1994.
13. Rani RJ, Chilambu CC. A survey on mining techniques for early lung cancer diagnoses, green computing, communication and conservation of energy (ICGCE). *International Conference* 2013; 918: 12-14.
14. Rajan JR, Chelvan CA. Mining gene expression using kohonen self organising map for lung cancer diagnoses. *International Journal of Applied Engineering Research (IJAER)* 2014; 9: 11883-11892.
15. Jing Li, Cheng JH, Shi JY, et al. Brief introduction of back propagation (BP) neural network algorithm and its improvement. *Advances in Computer Science and Information Engineering* 2012; 169: 553-558.
16. Freeman JA, Skapura DM. *Neural networks: Algorithm, applications and programming techniques.* 1999.
17. Howard DB. *Neural network toolbox for use with Matlab, The Mathworks Inc.* 4th ed. 2004.

### Correspondence to:

Juliet Rani Rajan,  
 Research Scholar,  
 Satyabhama University,  
 Infosys Technologies,  
 India.  
 Tel: 918939436673  
 E-mail: julietrajan@gmail.com