

# Asthma Prediction Using Classification Technique

J. Cathrin Princy\* and K. Sivaranjani\*\*

## ABSTRACT

Data mining is the method of analyzing data from similar perspective and summarizing it into useful information. Data mining is useful to find patterns to help in the important tasks of medical diagnosis and treatment. Asthma is one of the major cause of deaths and disabilities worldwide. In this paper development, variations in lifestyles, eating habits and population increase, these factors cause the number of asthma patients to increase year by year. In this paper proposed an integrated data mining system. In this system, two data mining methods named SVM and MLP are designed for predicting asthma attacks. The major methodology is to extract the significant information of asthma attacks and build classifier by using users' daily records and environmental data. Meanwhile, helpful medical information and suggestions supported by doctors are practical. In this way, the proposed system can predict the chances of asthma attacks and provide patients of the suitable medical instruction or health messages. The experimental evaluation results showed a 98.5% of accuracy and which proves this mechanism effective and reliable in asthma attack prediction.

**Keywords:** Data mining, Asthma, Breathing tests, SVM, MLP

## 1. INTRODUCTION

Data Mining is a knowledge mining process. It is an interdisciplinary subfield of computer science. It is the computational process of discovering patterns in big data sets involving methods at the intersection of artificial intelligence, machine learning, statistics and database systems. Because asthma is a chronic condition, it usually requires continuous medical care. Patients with reasonable to severe asthma have to take long-term medication daily (for example, anti-inflammatory drugs) to control the original inflammation and prevent symptoms and attacks. If symptoms occur, short-term medications (inhaled short acting beta2-agonists) are used to reduce them. Medication is not the only way to control asthma. It is also main to keep away from asthma triggers — stimuli that irritate and inflame the airways. Each person must study what triggers he or she should avoid. Although asthma does not kill on the scale of chronic obstructive pulmonary diseases (COPD), failure to use proper drugs or comply with treatment, coupled with an under-recognition of the severity of the difficulty, can lead to unnecessary deaths, most of which occur outside hospital. Asthma prediction physicians can make diagnostic decisions and treatment recommendations based on history, lab results and symptoms of patients. A analytical model can advise doctors whether a certain case would be better treated as an outpatient or an inpatient.

The neural networks may be regarded as the universal approximates of the measured data in the multidimensional space. They understand two types of approximation: the global and local one. The most important example of global network is the multilayer perception (MLP), employ the sigmoid activation function of neurons. In MLP the neurons are arranged in layers, including from the input layer (the set of input nodes), through the hidden layers, up to the output layer. The interconnections are permissible only between two nearest layers. The network is feed forward, i.e., the processing signal propagates from input to the output side. The most representative example of

\* Research Scholar, Department of Computer Science, Bishop Heber College, Tiruchirapalli, Tamilnadu, India, *Email: princycathrin@gmail.com.*

\*\* Asst. Professor, Department of Information Technology, Bishop Heber College, Tiruchirappalli, Tamilnadu, India, *Email: sivaranjani2000@yahoo.com.*

local neural network is the Support Vector Machine (SVM), of the Gaussian kernel function. The is a two layer neural network employing hidden layer of radial units and one production neuron. The process of creating this network and learning its parameters is organized in the method in which we deal only with kernel functions instead of direct processing of hidden unit signals.

This paper will summarize and compare these 2 networks: MLP and SVM. The comparison will be done with respect to the complexity of the construction as well as the accuracy of results for the solution of different learning tasks, including classification, prediction and regression problem.

## 2. RELATED WORK

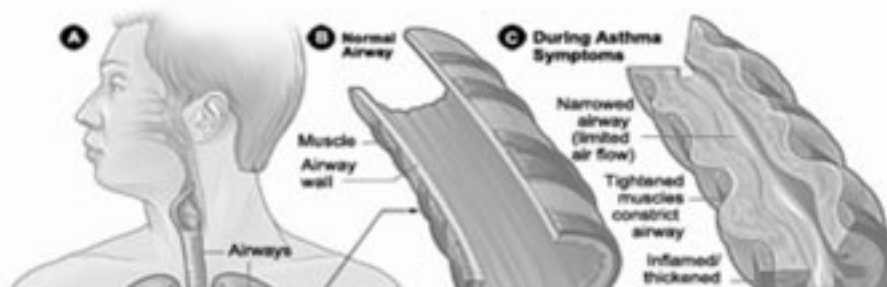
### 2.1. Asthma

Asthma (AZ-ma) is a chronic (long-term) lung infection that inflames and narrows the airways. Asthma causes recurring period of wheezing (a whistling sound as you breathe), chest tightness, shortness of breath, and coughing. The coughing commonly occur at night or early in the morning. Asthma affects people of all ages, but it most often starts throughout childhood. In the United States, more than 25 million populace are known to have asthma. About 7 million of these people are children.



#### 2.1.2. Overview

To understand asthma, it help to know how the airways effort. The airways are tubes that carry air into and out of your lungs. populace who have asthma have reddened airways. The inflammation makes the airways swollen and very sensitive. The airways be inclined to react strongly to certain inhaled substances. When the airways react, the muscles around them tighten. This narrow the airways, cause less air to flow into the lungs. The swelling also can worsen, making the airways flush narrower. Cell in the airways might make more mucus than usual. Mucus is a sticky, thick liquid that can further narrow the airways. This procession reaction can result in asthma symptoms. Symptoms can happen each time the airways are tenderness.



Asthma Figure A denotes the place of the lungs and airways in the body.

Figure B shows a cross-section of a ordinary airway. Figure C denotes the part of an airway during asthma symptoms.

Occasionally asthma symptom is mild and goes gone on their own or after minimal treatment with asthma medicine. Other times, symptoms continue to get worse.

### ***2.1.3. Signs and Symptoms of Asthma***

Ordinary signs and symptoms of asthma include:

*Coughing:* Coughing from asthma often is worse at night or early on in the morning, making it hard to sleep. *Wheezing:* Wheezing is a whistling or squeaky noise that occurs when you breathe. *Chest tightness:* This may feel like something is squeezing or sitting on your chest. *Smallness of breath:* Some people who have asthma say they can't catch their breath or they feel out of breath.

### ***2.1.4. Causes Asthma Symptoms to Occur***

Several things can trigger or get worse asthma symptoms. Triggers may include:

Allergens starting dust, animal fur, mold, and pollens from trees, grass, and plants. Irritants such as cigarette smoke, air effluence, chemical or dust in the workplace, compounds in home décor foodstuffs, and sprays (such as hairspray). Medicine such as aspirin or extra no steroidal anti-inflammatory medicine and nonselective beta-adrenergic blocking agent. Sulfates in food and drinks. Viral higher respiratory infections, such as colds. Physical activity, including exercise.

## **3. PROPOSED WORK**

### **3.1. Algorithms of SVM**

A Support Vector Machine (SVM) performs classification by finding the higher dimensional analogue of a plane that maximizes the margin between the 2 classes. The vectors (cases) that define higher dimensional analogue of a plane are the support vectors. In machine learning, support vector machines also support vector networks) are effective learning models with accomplice knowledge algorithms that study data used for classification and regression analysis.

#### ***3.1.1. Support Vector Machines: history I***

SVMs introduce in COLT-92 by Boser, Guyon & Vapnik. Become rather popular since. In theory well motivated algorithm: developed from Arithmetical Learning Theory (Vapnik & Chervonenkis) since the 60s. Empirically good show: successful applications in several fields (bioinformatics, text, image recognition)

#### ***3.1.2. Property of SVM***

Duality, Kernels, Margin Convexity, Sparseness

#### ***3.1.3. SVM Application***

- Linear SVM Classifiers
- Nonlinear SVM Classifiers
- Solution of SVM Formulation
- SVM Kernel Functions

#### ***3.1.4. Advantages of Support Vector Machines***

Effective in high spatial spaces. Still effective in cases where number of dimension is larger than the number of samples. It is used the decision function that has subset of training points (called support vectors), so it is also

memory resourceful. Resourceful: various kind of Kernel functions can be specified for the choice function. Ordinary kernels are provided, but it is also possible to specify custom kernels.

### 3.1.5. Classification SVM Type 1

For this category of SVM, training includes the error decreasing function:

$$\frac{1}{2} w^T w + C \sum_{i=1}^N \zeta_i$$

Subject to the constraint:

$$y_i w^T \phi(x_i) + b \geq 1 - \zeta_i \text{ and } \zeta_i \geq 0, i = 1, \dots, N$$

Where C is the capacity invariable, w is the coefficient vector, b denotes the constant, and  $\zeta_i$  represents parameters for managing unseparable data (inputs). The index denoted by i label and the training cases are denoted by N. Note that  $y \in \pm 1$  represents the class label and  $x_i$  represent the separate variables. The kernel function  $\phi$  is used to transform data from the input (independent) to the attribute space. It should be noted that the bigger the C, the more the error is penalized. Thus, C should be chosen with care to avoid over fitting.

### 3.1.6. Classification SVM Type 2

In contrast to Cataloging SVM Type 1, the Classification SVM Type 2 model decreases the error function:

Subject to the constraints:

$$y_i w^T \phi(x_i) + b \geq \rho - \zeta_i, \zeta_i \geq 0, i = 1, \dots, N \text{ and } \rho \geq 0$$

In a regression SVM, you have to estimate the functional dependence of the reliant variable y on a set of self-determining variables x. It assumes, like other regression problems, that the relationship among the independent and needy variables is known by a deterministic function f plus the addition of some additive noise:

## 3.2. Algorithms of Multilayer perception

It is a feed forward artificial neural network model that maps set of input data onto a set of suitable outputs. An Multilayer perception consists of multiple layers of nodes in a directed graph, with every layer fully connected to the after that one. Apart from for the input nodes, each node is a processing element with a non-linear creation function. MLP use a supervised learning technique called backpropagation for training network. MLP is a alteration of the normal linear perception and can distinguish data that are not linearly separable. The learning process of MLP network is based on the information samples, composed of the x that is N-dimensional input vector and d is the M-dimensional preferred output vector, called objective. By processing the input vector x the MLP produces the production signal vector  $y(x, w)$ , where w is the vector of made to order weights. The error signal produced actuate a organize mechanism of the learning algorithm. The curative adjustments are designed to create the output signal  $y_k (k = 1, 2 \dots M)$  to the valid response  $d_k$  in a step by step manner.

The learning algorithm of MLP is based on decreasing the error function defined on the learning set  $(x_i, d_i)$  for  $i = 1, 2 \dots p$  using an Euclidean norm

$$E(w) = \frac{1}{2} \sum_{i=1}^p \|x_i, w - d_i\|^2$$

The minimization of this error leads to the optimal values of weights. The most efficient method of decreasing is the gradient algorithms, from which the most effective is the Levenberg Marquardt algorithm for middle volume networks and consolidate gradient for large size networks. Generally in all gradient algorithms the edition of weights are performed one by one process according to the following scheme

$$w(k+1) = w(k) + \eta \Delta w(k)$$

In Levenberg-Marquardt approach the least square formulation of knowledge problem is exploited and solved by using second order method of Newton type.

$$E(w) = 0.5 \sum_{i=1}^M (w - d_i)^2$$

#### 4. EXPERIMENT AND RESULTS

The proposed system has been implemented using .NET environment. It can be executed on windows. The results are obtained as follows after classified ads. Sample data set is presented. The Main theme of this literature survey is asthma prediction of people. Data has been collected from various resources like hospitals. This measurement describes the amount of air excluded from the lungs during the middle half of the forced vital capacity test. Many physicians like to look at this worth because it is an indicator of obstructive disease. In this paper two types of algorithms are used to predict whether the person has asthma or not. The algorithms are namely MLP and SVM. This paper, our goal is how to easily predict asthma affected people and then give more accuracy compare than other classification technique. This research utilize the data set contain the electronic medical details of different patients. This include patient's name, infection age, sex, date, address, etc, in particular year. Fig. 1. shows the bar graph of the number of no of diseases, asthma diseases affecting the patients monthly. Fig. 2. Depicts number of patients affected by asthma diseases month. It unfold the fact that in a particular month some patients are affected by the same disease.

In this study, various classification techniques are used to find out the sensitivity, specificity, accuracy of asthma prediction data. High accuracy achieved through SVM technique compare than MLP.

**Table 1**  
Shows the Table of the number of diseases affecting the patient's dataset

| S. No. | Disease | Condition |
|--------|---------|-----------|
| 1      | Cough   | Asthma    |
| 2      | Phlegm  | Others    |
| 3      | Wheeze  | Asthma    |
| 4      | Smoking | Asthma    |
| 5      | Fever   | Others    |



**Figure 1: Sensitivity, Specificity, Accuracy, Time found in SVM**



Figure 2: Sensitivity, Specificity, Accuracy, Time found in MLP

#### 4.1. Comparison of Various Dataset using Algorithms

Indicating the sensitivity, specificity accuracy, time taken of the 2 methods used in this study is given in Table 2.

Table 2

| Predict-or | Sensitivity | Specificity | Accuracy | TimeSec |
|------------|-------------|-------------|----------|---------|
| SVM        | 97.5%       | 98.9%       | 98.5%    | 0.07    |
| MLP        | 93.5%       | 92.9%       | 90.5%    | 0.10    |

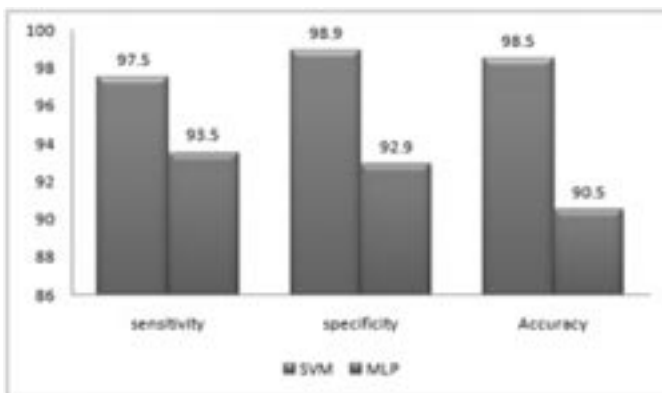


Chart 1: Compare to SVM & MLP sensitivity, Specificity, accuracy highest in svm

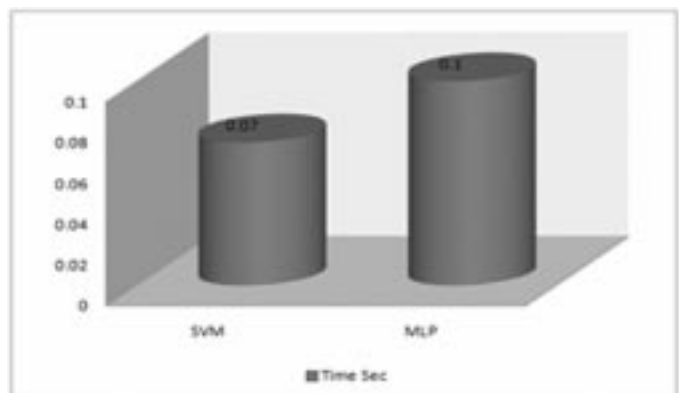


Chart 2: Compare to SVM & MLP time second fast in SVM

## 5. CONCLUSION

In this study, classification techniques are used to find out the sensitivity, specificity, accuracy, time second of asthma prediction data. High accuracy achieved through SVM technique compare than MLP. SVM achieved 97.50% sensitivity, 98.90% specificity, accuracy 98.50% compare than other algorithms. Three types of breathing tests are use to find out asthma prediction which are FEV1, FVC and FEF. In this paper, commonly applying the FEV1/FVC test to find out the lung capacity. Lung capacity means find out the various breathing sounds of a particular person. Thus the conclusion of this paper produces the result of getting more accuracy through the comparison of SVM and MLP technique. Finally, SVM technique is best compare than other classification technique.

## REFERENCES

- [1] Ali, J, summer, WR, & Levitzky, MG. "Pulmonary pathophysiology" (3rd ed.). New York: *McGraw-Hill Medical*, 2009.
- [2] BDCN Prasad, P E S N Krishna Prasad and Y Sagar -"A Comparative Study of Machine Learning Algorithms as Expert Systems in Medical Diagnosis (Asthma)", *CCSIT 2011, Part I, CCIS 131*, 570– 576, 2010.
- [3] BDCN Prasad, P E S N Krishna Prasad, Y Sagar: "A Study on Associative neural memories", *IJACSA*, **1**, 2010.
- [4] BDCN Prasad, P. E. S. N Krishna Prasad and Y Sagar, "An approach to develop expert systems in medical diagnosis using machine learning algorithms (Asthma) and a performance study", *International Journal on Soft Computing*, **2(1)**, 26-33, 2011.
- [5] A. Bush, "Diagnosis of asthma in children under five," *Primary Care Respiratory Journal*, **16(1)**, 7–15, 2007.
- [6] Çatmaka°, Z, Köse, Y, Toker, O, & Öz, HR. "Towards an ARM based low cost and mobile biomedical device test bed for improved multi-channel pulmonary diagnosis". In 4th European Conference of the International Federation for Medical and Biological Engineering, *Berlin Heidelberg: Springer*, 1108–1112, 2009.
- [7] E. Chatzimichail, E. Parakakis and A. Rigas, "Predicting asthma outcome using partial least square regression and artificial neural networks" *Advances in Artificial Intelligence*, Article ID 435321, 1-7, 2013.
- [8] E. Chatzimichail, E. Parakakis, M. Sitzimi and A. Rigas, "An intelligent system approach for Asthma prediction in symptomatic preschool children", *Computational and Mathematical Methods in Medicine*, Article ID 240182, 1-5, 2013.
- [9] M Gori, P Frasconi and A Tesi, "Success and failures of Backpropagation: A Theoretical Investigation".
- [10] J.R.G.Lansangan and E.B.Barriors, "Principal components analysis of nonstationary time series data", *Statistics and Computing*, **19(2)**, 173-187, 2009.
- [11] Nahit Emanet, Halil R Oz, Nazan Bayram and Dursun Delen, "A comparative analysis of machine learning methods for classification type decision problems in healthcare", *Decision Analytics*, **1(6)**, 1-20, 2014.
- [12] TahaSamadSoltaniHeris, MostafaLangarizadeh, Zahra Mahmoodvand, Maryam Zolnoori, "Intelligent diagnosis of Asthma using machine learning algorithms", *International Research Journal of Applied and Basic Sciences*, **5(1)** , 140 – 145, 2013.
- [13] Xie J, Wang C."Using Support vector machines with a novel hybrid feature selection method for diagnosis of erythematous diseases. " *Expert Syst. Appl.* **38(5)**, 5809-5815, 2011.