Spirometry Data Classification Using Self Organizing Feature Map Algorithm

[1]Kamlesh Waghmare and [2]Dr. P. N. Chatur

[1]Asst.Professor, Department of CSE, PRMIT& R Badnera, Amravati, Maharashtra, India.

waghmare.kamlesh@gmail.com

[2]HOD, Department of CSE, GCOE, Amravati, Maharashtra, India.

Abstract

In this work the classification of Force Expiratory volume in 1 second (FEV 1) in pulmonary function test is carried out using Spirometer and Self Organizing Feature Map Algorithm. Spirometry data are measure with flow volume spirometer from subject (N=100 including Noramal, and Abnormal) using standard data acquisition protocol. The acquire data are then used to classify FEV1. Self Organizing Map was used to classify the values of FEV1 into Normal, Obstructive and Restrictive. The Spirometry data was statistically analyzed for neural network. The FEV1 parameters were presented as inputs to Self Organizing map algorithm. The self organize map classified normal and abnormal classes, abnormal class again classified into Obstructive and restrictive classes. The result shows the Accuracy, Sensitivity and Specificity of Self Organizing Map algorithm.

Keywords: Force Expiratory volume in 1 second (FEV1), Neural Network, Self Organizing Feature Map (SOM), Spirometer.

I. Introduction

Spirometer is a relatively simple, noninvasive and the most widely used pulmonary function test that measures the volume of air expelled from fully inflated lungs as a function of time. Spirometer measures the volume of air inhaled or exhaled as a function of time during forced breathing maneuvers and is an essential tool in the diagnosis of airway obstruction and the detection of respiratory diseases [1].

Accurate measurement of respiratory function is necessary to assess and manage different pulmonary abnormalities.The following parameters are measured using Spirometer such as Vital Capacity (VC), Forced Vital Capacity (FVC), Forced Expiratory Volume at 1st second (FEV1), Peak Expiratory Flow (PEF) and Forced Expiratory Flow 25-75% (FEF 25-75) [2]. In this work, the ratio of FEV1 to FVC was computed known as FEV1% and use as input to Self Org. Spirometer parameter depends on factors such as age, height, weight, and gender of subjects.

Obstructive and restrictive diseases are two main types of lung disease found with lung function tests. In obstructive lung conditions the airways are narrowed usually causing an increase in the time it takes to empty the lungs. Obstructive lung disease can be caused by conditions such as emphysema, bronchitis, infection (which produces inflammation) and asthma [3], [4]. Neural Networks have been a natural choice as trainable pattern classifiers because of their capability to approximate functions and to generalize. Artificial Neural networks (ANN) are powerful computational systems consisting of many simple processing elements connected together to perform tasks analogously to biological brains. They are massively parallel, which makes them efficient, robust, fault tolerant and noise independent. They can learn from training data and generalize them to new situations. The learning process of the ANN is similar to the learning function of the brain. During training, samples are presented to the input layer that yields changes of the activation state of output processing elements [5]. The calculated output value is compared to the required value which is also given in the training set.

.

II. Related Work

Many researchers have tried to use Artificial Neural Network and Data Mining technologies in areas related to Spirometry data classification.

Mahdi Jan et.al. used the Multilayer Perceptron Neuaral Network (MLPNN) for detection of normal and restrictive pattern of pulmonary disease. In that study they took all three patterns of respiratory abnormalities into consideration. For the classification they used recurrent neural network. The accuracy, sensitivity and specificity of that algorithm were also calculated.[5]

Anandan K worked on the diagnostic of relevance of Spirometric pulmonary function test using Neural Networks and Principal Component Analysis [PCA]. Principal Component Analysis was performed on the data sets with measured and predicted values. PCA transforms the input space into a new lower dimensional space. High accuracy was measured of PCA and ANN. [6]

Kavitha A. worked on the classification and prediction of Spirometry data using Support Vector Regression Algorithm. In this, they classify Spirometry data into normal and abnormal using Regression Algorithm and accuracy was measured.[7]

Chatur P.N. et al used ANOVA technique and SPline function for the classification of Spirometry data. In this study they used Spirometry is in the form of graphs i.e., flow-volume loop and volume-time curve for increase the efficiency of algorithm.

III. Methodology

There are different steps in implemented and various methodology used in implementation. Breeze software is use to connect Spirometer device with PC and Matlab 7.0 is use. Spirometry data of 100 adult volunteers (50 Normal, 30 Restrictive, 20 Obstructive) are use in this work. The data is divided to training and test sets. The age, gender, and other parameter of patients are identified and use for obtaining dataset of FVC, FEV1. For Spirometry data classification uses the Self Organizing Feature Map Algorithm. Self Organizing Feature map is used for unsupervised classification and prediction. The Self-Organizing map is comprised of a collection of codebook vectors connected together in a topological arrangement, typically a one dimensional line or a two dimensional grid. The Best Matching Unit (BMU) is the codebook vector from the pool that has the minimum distance to an input vector. A distance measure between input patterns must be defined by Euclidean distance [6].

Dist(x,c) = sum(i=1)n (x i – c i)^2……………………………1

Where n is the number of attributes, x is the input vector and c is a given query vector

ci(t+1) = learn_rate(t) \times (ci(t) – xi) …………………..2

where c i(t) is the ith attribute of a query vector at time t, learn_{rate} is the current learning rate, an x i is the ith
attributes of a input vector.

IV. Dataflow Diagram

The four main stages of the project are as follows:

1. Data Collection.
2. Data Preprocessing.
3. Artificial Neural Network Technique
4. Data Classification
5. Calculate Accuracy and Sensitivity

a. Data Collection

The most important part while implementing any data related project is collection of proper data for the analysis .Thus, in this work acquires spirometry data from Spirometer device for classification.

b. Data Preprocessing

Important step in this work was process a large data preprocessing. One of the challenges that face the Neural Network process in Spirometry data is poor data quality .For this reason we try to prepare our data carefully to obtain accurate and correct Title and Authors

c. Artificial Neural Network

A Neural Network model is a structure that can be adjusted to produce a mapping from a given set of data. The key element of biological Neural Network paradigm is the novel structure of the information processing system. It is composed of a large An Artificial Neural Network (ANN). Artificial Neural Network is an information processing paradigm that inspired by number of highly interconnected processing neurons. An ANN is configured for a specific application, such as pattern recognition or data classification, through a learning process. Learning in biological systems involves adjustments to the synaptic connections that exist between the neurons.

There are two types of Neural Networks [7]

I. Supervised
II. UnSupervised

I) Supervise Learning

In Supervised learning, a desired output result for each input vector is required when the network is trained. An ANN of the supervised learning type, such as the multi-layer perceptron, uses the target result to guide the formation of the neural parameters. It is thus possible to make the Neural Network learn the behavior of the process under study.

II) Unsupervised Learning

In Unsupervised learning, the training of the network is entirely data-driven and no target results for the input data vectors are provided. An ANN of the unsupervised learning type, such as the Self-Organizing Feature map, can be used for clustering the input data and find features inherent to the problem. Adaptation of the weight vectors of a neuron occurs through a similar process to competitive learning except that subsets of nodes are adapted at each learning step in order to produce topologically ordered maps. This also means that the weight vectors on the neurons adapt so as to become ordered along the axes of the network

d. Data Classification

In data classification step the subject is classified according to age, FVC, FEV, FEV1, Weight and Height into Normal and Abnormal. Abnormal Data again train and classified into Restrictive and Obstructive.

e. Calculate Sensitivity and Accuracy

The Following Formulas are used for calculation Accuracy and Sensitivity

$$Accuracy = (TP+TN)/ (TP+FP+TN+FN) \quad 3$$

$$Sensitivity = TP / (TP+FN) \quad 4$$

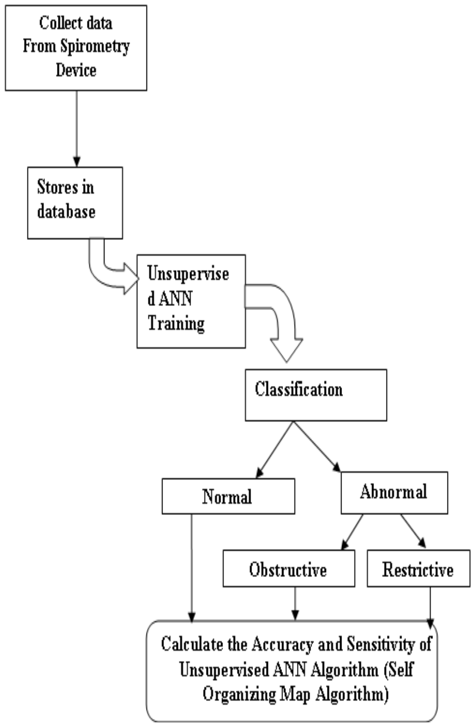

Fig 1.Classification Model

V. Result and Discussion

Artificial Neural Networks are appropriate alternatives to standard statistical methods. In order to train a Neural Network model appropriately, the samples must be selected carefully in order to represent the whole subjects under investigation. In this work, 100 samples were used to train the network and test it. The following figure shows the classification of classification of Spirometry data according to FVC/FEV ratio. It classifies the data into normal and abnormal class and accuracy 93% is measured.

The performance of the neural networks was estimated using false positive (FP), False Negative (FN), True Positive (TP) and True Negative (TN) value Classification of a normal data as abnormal is considered as FP and classification of abnormal data as normal is considered FN. TP and TN are the cases where the abnormal is classified as abnormal and normal classified as normal respectively

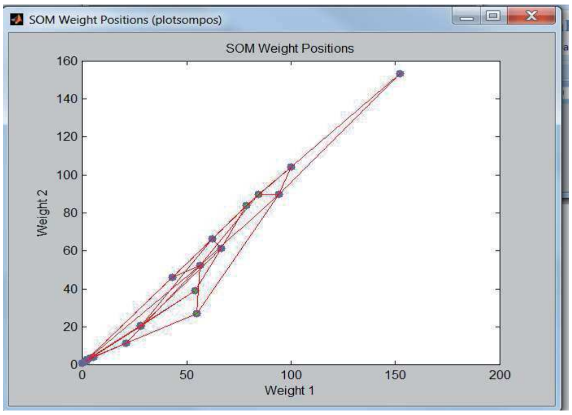

Figure 2. SOM Weight position

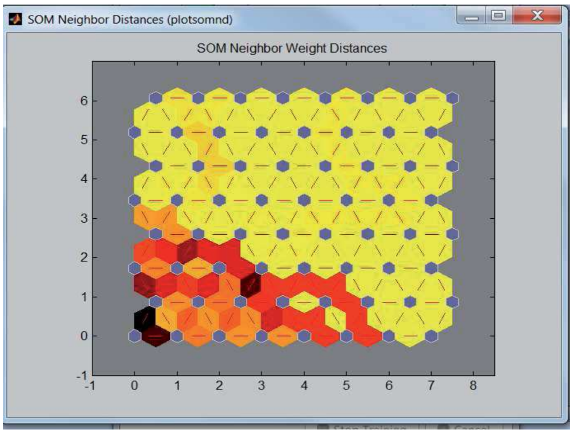

Figure.3. Neighbor weight distance

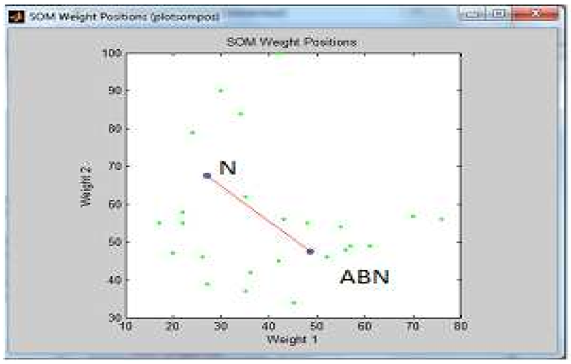

Figure 4 Classifications of Data into Normal and Abnormal

N-Normal
ABN-Abnormal
Above figure shows the Classification of data into Normal and Abnormal Class. The bottom point shows the Normal Class and Top point shows the Abnormal Class.

VI. Conclusion

The Spirometer has been comprehensively used for diagnosis of lung disease.Various Supervise Artificial Neural Network algorithms have been proposed in literature for Spirometry data classification. The parameters like FEV1, FVC, PEV and PEF are classified according to the test data. Many researchers proposed different methods for classification methods. Self
Organizing Feature Map used two dimensional topological structures for classification. It is well suited for one or more dimensional classes. This method aims to classify and predict Spirometer parameter such as Force Vital Capacity, Force Expiratory Volume in one second, and Small Vital Capacity into Normal, Obstructive and Restrictive. It shows that ANN can be used as an alternative to other statistical methods. The Accuracy of Self Organizing Feature Map 93% is measure.

VII. References

[1]. A. Kavitha1, C. M. Sujatha and S. Ramakrishnan,”*Prediction of Spirometric Forced Expiratory Volume (FEV1) Data Using Support Vector Regression*” MEASUREMENT SCIENCE REVIEW, Volume 10, No. 2, 2009

[2]. Crapo, M.D., 1994. Pulmonary-function testing. N. Engl. J. Med., 331: 25-30

[3]. Oud, M. and Maarsingh, E.J.W. (2004). Spirometer and forced oscillatory assisted optimal frequency band determination for the computerized analysis of tracheal lung sounds in asthma. Physiol. Meas. 25, 595-606

[4]. T.Kalpalatha and N.Kumaravel “Texture Analysis of bone CT Images for Classification and Characterization of bone Quality”International journal of Soft Computing 4(5):223-228,2009

[5]. K.A.Waghmare,”Spirometry data Analysis and Classification Using Artificial Neural Network:IJETAE,ISSN 2250- 2459,Vol.2 Issue 1 2012

[6]. Lim Joalin P.K. et al “Pulmonary function measurement using flow-volume “0-7803-5154-9/98/IEEE

[7]. Sujatha C. Manoharan,”Comparison of Two ANN Methods for Classification of Spirometry Data”, Measurement ScienceReview,Vol(8),Section 2,No.3,2008.