Available at: www.researchpublications.org

Spirometry Data Analysis and Classification Using Data Mining: An Approach

[1]K. A. Waghmare, , [2]Dr. G.R. Bamnote, [3]Dr. P. N. Chatur [4]M.B.Waghmare
[1] *Department Computer Science & Engineering, P.R.M.I.T& R, Badnera,*
[2]HOD, Department Computer Science & Engineering, P.R.M.I.T& R, Badnera,
[3]HOD, Department Computer Science & Engineering,GCOE,Amravati.
[4] Department Computer Science & Engineering, P.R.M.I.T& R, Badnera,
waghmare.kamlesh@gmail.com
grbamnote@rediffmail.com
chatur.prashant@gcoea.ac.in
milind.btk gmail.com

Abstract- In this paper, the acquisition of Spirometry data such as Forced Expiratory Volume in 1 second (FEV1), Force Vital Capacity (FVC) and Small Vital Capacity will be carried out using Spirometer. At current, numbers of lung diseases are main hazard to the human health due to air pollution, smoking and other infections. The various Artificial Neural Network methods for the taxonomy of Spirometry data are Back Propagation Network (BPN), Radial Basis Function (RBF) and Multilayer Perceptron Neural Network (MLPNN). The aim of the present study is to acquire parameter such as FVC, FEV1 and SVC data and use Data Mining Algorithm for classification of the Spirometer data into Normal, Obstructive and Restrictive dataset. This approach is used to increase the efficiency of classification

Keywords – Data Mining, Spirometry, Obstructive, Neural Network, Multi Layer Perceptron Neural Network, Force Expiratory Volume in 1 second, Back Propagation Network,Restrictive.

I. INTRODUCTION

Nowadays respiratory diseases are one of the most serious problems for the human health. Respiratory diseases are curable from early detection. Respiratory diseases are mostly caused due to factors such as air pollution, smoking and various infections. Respiratory function is commonly access by the standard Spirometry pulmonary function test. There are two main type of lung diseases found with lung function test such as Obstructive and Restrictive [1]. In Obstructive lung condition the airways are narrow normally causing an increasing in the time taken to empty lung [2]. Spirometer device measures the volume of air inhaled or exhaled as a function of time during the breathing. It is the most important tool in the diagnosis of airway obstruction disease [3].

Spirometer test depends on some factor such as age, height, weight, gender, and location i.e. in which location the subject are living for example the people living in hilly areas have more breathing capacity as compare to other people. The Spirometer also used to classify the respiratory disease from cardiac disease. Obstructive and restrictive pattern have four different levels as categories as follows [4].

1. Mild
2. Moderate
3. Severely Moderate
4. Severe

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]

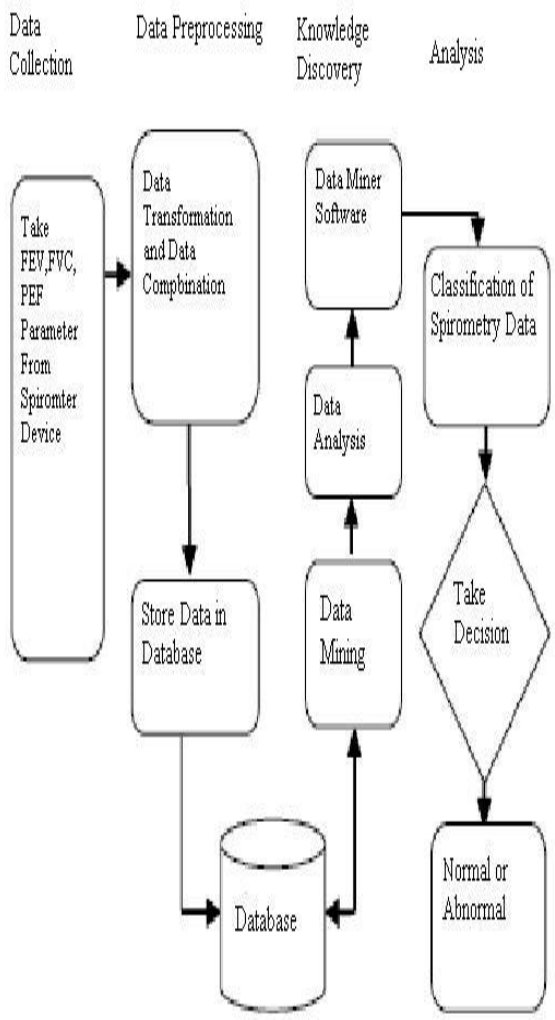
Kavitha A. worked on the classification and prediction of Spirometry data using Support Vector

Regression Algorithm. In this, they classify Spirometry data into normal and abnormal using Regression Algorithm and accuracy was measured. [7]

Chatur P.N. 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. [8]

III. PROPOSED METHODOLOGY

This work is completely designed in four different stages as shown in the fig. below:



The four main stages of the project are as follows:
1. Data Collection.
2. Data Preprocessing.
3. Knowledge Discovery.
4. Data Mining Technique.

A. Data Collection

The most important part while implementing any data related project is collection of proper data for the analysis using any technique for (e.g. Data Mining).Thus, In this work acquire spirometry data from Spirometer device
for classification.

B. Data Preprocessing

An important step in the data mining process is data preprocessing. One of the challenges that face the knowledge discovery process in climate data is poor data quality. For this reason we try to prepare our data carefully to obtain accurate and correct results.

C. Knowledge discovery

For knowledge extraction various data mining techniques are available such as Outlier Analysis, Clustering, Prediction and Classification and Association Rules. In this work, use Classification technique of data mining for our purpose, because Classification has been utilized to predict patient lung condition.

D. Data Mining Technique

Technologies have been elaborated over the last few years, producing a huge amount of data. This huge raw data is difficult to analyze and understand. In this study, clustering aim to improve the understanding of Spirometry data.[9]
In this work k-means clustering algorithm will be use for classification. Using k=6. K-means algorithm is the most popular clustering tool used in scientific and medical applications. The name comes from representing each of k clusters by the mean (or weighted average) of its points, the so-called centroid. The centroid of a cluster is a point whose coordinates are the mean of the coordinates of all the points in the clusters. Prediction is the most used data mining task in the field of meteorology. Data mining techniques provides with a level of confidence about the predicted solutions in terms of the consistency of prediction and in terms of the frequency of correct predictions. [10][11]

IV. FUTURE SCOPE

In Spirometry data Classification there are various Supervise Artificial Neural Network Algorithms and Data Mining were used. In future the Statistical Data miner and K-Means Clustering Algorithm can be use to predict whether subject is in normal state or abnormal state, it can also be use to implement for more accurate result.

V. CONCLUSION

Thus in this paper, the parameters like FEV1, FVC, FEC and MVV will be classified according to the test data. K-Means clustering method is easy to implement. Many researchers used different methods for classification methods. K-Means clustering Algorithm is use for classification. It is well suited for one or more dimensional classes .thus, the proposed method aims at classify and predict Spirometer parameter such as Force Vital Capacity, Force Expiratory Volume in 1second, and Small Vital Capacity into Normal, Obstructive and Restrictive. This approach is used to increase the efficiency of classification

REFERENCES

[1]. Elman J. 1990. "*Finding Structure In Time*", Cognitive Since. 14: 179-211. DOI: 10.1016/0364- 0213(90)90002-E.

[2]. Abboud, S. et. Al., "*Maximum expiratory flow-volume curve: mathematical model and experimental results*". Med. Eng. Phys. 17 (5), 332-336

[3]. Gaetano P. and Marieann H. "*Assessment of respiratory mechanism by artificial neural network: an exploratory study*". *J. Appl. Physiol*. 90 (5), 1817-1824

[4]. Waghmare K.A.et. al."*Analysis and classification using Artificial Neural Network*", IJETAE, vol. 2, Issue, Jan-2012.

[5]. Mahdi Jan Baemani et.al."*Detection of Respiratory Abnormalities Using Artificial Neural Networks*", Journal of Computer Science 4 (8): 663-667.

[6]. Anadan K. et. Al.," *Evaluation of Flow- Volume Spirometric test Neural Network Based prediction and Principal Component Analysis*",,J.Med SYST[2011] 35:127-133 DOI: 10.1007/S10916- 009-9349-7

[7]. Kavitha A. et. Al., "*Prediction of spirometric Forced Expiratory Vollume data using Support Vector Regression*", MEASUREMENT SCIENCE REVIEW, Volume 10, No. 2, 2010.

[8]. Chatur P.N.,"*Classification of pattern Using Spline Function and Statistical Data mining*", Journal Of Computer Technology & Applications, VOL 3, NO 1 (2012)

[9]. Badhiye S.S." *Temperature and Humidity Analysis using Data Logger of Data Acquisition System*" IJETAE, vol. 2, Issue, Jan-2012.

[10]. Baboo S., and Shereef K., "*Applicability of Data Mining Techniques for Climate Prediction – A Survey Approach*," International Journal of Computer Science and Information Security, Vol. 8, No. 1, April 2010.

[11]. Bilgin T., and Çamurcu Y., "*A Data Mining Application on Air Temperature Database,*" Advances in Information Systems, Springer Berlin, Heidelberg, pp.68-76 .2004.