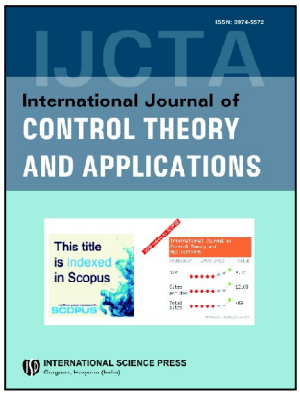

International Journal of Control Theory and Applications

ISSN : 0974-5572

Volume 10 • Number 10 • 2017

Analysis of Data Mining Techniques used in Spirometry Data

[1]K. Rohini and [2]G. Suseendran

[1]Ph.D Research Scholar, Department of Information Technology, School of Computing Sciences, Vels University, Chennai, India, E-mail: rohinik1512@gmail.com

[2]Assistant Professor, Department of Information Technology, School of Computing Sciences, Vels University, Chennai, India

Abstract: The objective of this paper to the analysis of recent research work is done on various data mining techniques, methods and methodologies used in effective data analysis under medical domain especially in analyzing spirometry data which is obtained from pulmonary function tests (PFT). Pulmonary function tests is used to identify lung capacity, based on which the various lung diseases can be identified. This paper examines the prospective application of data mining techniques such as Associative Rule based, Naïve Bayes, Decision tree and Artificial Neural Network for the classification of enormous amount of medical domain data. It pays attention on survey and review of data mining techniques used for analyzing the factors of lung disease and current spirometry data classification to discover hidden information. Spiro meter plays a significant role in diagnosing most of the lung diseases. Recently numerous researches and techniques have developed classification and prediction of Spirometry data which is used to predict the lung capacity. The proposed schemes were mostly based on Artificial Neural Network. The function of respiratory system based on the ventilation mechanism. The quality of lung depends upon the quality of lung ventilation. Findings: As a result the purpose of this paper is to give an overview of recent publications concerning the application of data mining techniques on medical domain, especially on lung diseases using spirometry data. The various Artificial Neural Network techniques were discussed for classifying normal and abnormal spirometry data. In addition, this paper also provides comparative study of various methods proposed by researchers.

Keywords: Spirometry, Obstructive, Restrictive, Respiratory disease, Pulmonary, Bronchus, Pulmonary function test, Data Mining, Asthma, Asscocitative rule.

1. INTRODUCTION

Air pollution, smoking and various infections are the most influencing factors for many of the respiratory diseases. Early and correct detection of respiratory problems is necessary in many associated treatments[1]. Measurement of respiratory function is essential to detect various pulmonary abnormalities. A variety of techniques are used to identify pulmonary problem such as pulmonary function tests (PFT). Pulmonary function tests are a series of tests which determines the lung capacity during inhalation and exhalation and it also finds the quantity of oxygen which is taken from the atmosphere into the body for circulation. Respiratory function is usually accessed by the normal pulmonary function test using spirometry device. Obstructive and Restrictive are the two main types of lung abnormalities recognized with pulmonary function test[2]. The bronchus is an airway in the respiratory tract

that conducts air into the lungs. In Obstructive pulmonary condition, the bronchus gets narrowed down causing difficulty and delay in breathing in and out[3]. Rigidness of the chest wall, weakness of muscles and damaged pulmonary nerves are all the factors that cause restriction in lung expansion. Restrictive pulmonary disease most often results from a condition causing stiffness in the lungs, that restrict lung expansion, resulting in a decreased lung volume, an increased effort of breathing, and inadequate ventilation and oxygenation[4].Spirometry test is used to assess lungs work by measuring the volume of air inhaled or exhaled and how quickly it is exhaled[5].The Spirometry test data are inferred depending up on many other factors such as age, height, weight, gender, location and smoking habit. The Spirometry test data depicts the lung functionalities during breathing in and breathing out. Spirometry test is used to diagnose asthma, chronic obstructive pulmonary disease (COPD) and other conditions like pulmonary fibrosis and cystic fibrosis that affect breathing[6].

The following are the terminologies used in pulmonary function tests:

- FVC is the abbreviation of Forced vital capacity. It is the total volume that can be vigorously exhaled from a maximum inhalation effort (L)
- FET& FIT —Forced expiratory time, which gives the amount of time the patient exhales through the FVC progress (seconds), forced inspiratory time.
- ERV is Expiratory reserve volume measures the highest volume of air that can be exhaled from the end-expiratory tidal position (L).
- FEV1 — Forced expiratory volume in 1 second; volume of air forcibly expired from a utmost inspired effort in the first second (L).
- FEV1/FVC ratio — Ratio between FEV1 and FVC.
- FRC — Functional residual capacity; the quantity of air in the lungs subsequent to a tidal volume exhalation = ERV + RV (L).
- PEF (peak expiratory flow). PEF is the highest exhalation flow rate attained

The two different curves obtained in the spirometry tests are flow-volume loop and volume-time curve. The flow-volume loop compares the changes in both the flow rates and the volume formed at different points of the FVC and FIV maneuver. The pulmonary function test primarily depends on some parameters such as Force Vital Capacity, Force Expiratory Volume1, Peak Expiratory Volume and Small Vital Capacity. The environmental parameter has greater influence in allergic hypersensitivity which induces the symptoms of asthma. Environmental Atopy refers to an allergic hypersensitivity that affects different parts of the body that do not come in contact with allergens, substances that trigger the body's allergic reaction. Atopy can include eczema (atopic dermatitis), allergic rhinitis, allergic conjunctivitis, and asthma. A number of children with eczema or atopic dermatitis extend to asthma. Several findings show that children with atopic dermatitis may have more severe and persistent asthma like adults. The children with family history of asthma have greater influence in childhood asthma. The inherited genetic makeup predisposes to have asthma. According to a CDC report, the children with the family history of asthma are more prone to asthma than who does not have a parent with asthma. In this paper various Artificial Neural Network techniques are discussed for classifying normal and abnormal Spirometry data. In addition, this paper also provides comparative study of various methods proposed by researchers. Many researchers have done their work in this area related to Spirometry data classification using techniques of Data Mining and Artificial Neural Network.

Anadan. K [10] analyzed pulmonary function test data using Neural Networks and Principal Component Analysis [PCA]. He measured and predicted data values which are analyzed with Principal Component Analysis. PCA is also used to transform the input space into a new lower dimensional space and for high accuracy measure. Jafari S, Arabalibeik H, [8] agin focused to propose a system to detect functionalities of pulmonary system by

using spirometry data and multilayer perception neural networks (MLPNN). The respiratory patterns are detected and categorized into normal, obstructive and restrictive. Baemani M H [7] proposed the Multilayer Perception Neural Network (MLPNN) and recurrent neural network for detection and classification of the entire three patterns of respiratory abnormalities.Mahesha V, Ramakrishnana.S [6] emphasized their work, to classify spirometry data with feed forward neural networks and to generate flow – volume curves. The pressure and resistance parameters are obtained from activation function representing the pressure – volume association of the lung. During maximum expiration the curves of pressure – time and resistance – expiration volume were obtained. Kavitha A [9] proposed Support Vector to classify and predict Spirometry data. Regression Algorithm. Regression Algorithm was used to categorize abnormalities of Spirometry data. Zhanqi Zhao, Knut Möller Barbara Vogt, Inéz Frerichs[11] reported an example of customized software for one of these clinical trials and demonstrate the development of the software for batching EIT data. Chathuri Daluwatte', Christopher [13] presents a promising approach to detect breathing peaks in respiration signals from spontaneously breathing subjects. The algorithm was able to identify breathing peaks consistently while the breathing rate varied from 15 to 160 breaths/min and also in presence of noise and motion artifacts.

2. PROPOSED METHODOLOGY

The proposed Analysis is as follows: Descriptive analytics which answers the questions what happened and why did it happen. Descriptive analytics looks at past performance and understands that performance by mining historical data to look for the reasons behind past success or failure. Prescriptive analytics can continually take in new data to re-predict and re-prescribe, thus automatically improving prediction accuracy and prescribing better decision options. Predictive analytics encompasses a different category of statistical techniques from predictive modeling, machine learning and data mining that analyze present and historical facts to make predictions about future or unknown events. Prescriptive analytics is a combination of data, mathematical models and various business rules. The data for prescriptive analytics can be both internal (within the organization) and external (like social media data). Business rules are preferences, best practices, boundaries and other constraints. Mathematical models include natural language processing, machine learning, statistics, operations research, etc. Data mining techniques square measure is widely used for prognostic analysis. The method involves extracting data from an information set associate degreed convert it into an affordable formation for more use. Here we have a tendency to use data processing and neural networks strategies and methodologies to spirometry knowledge analysis, detection and classification. Two curves gift within the spirometry tests that square measure flow-volume loop and volume-time curve. The flow-volume loop compares all the flow rates and therefore the volume changes shaped at totally different points of the FVC and FIV maneuver in Figure 2.1.

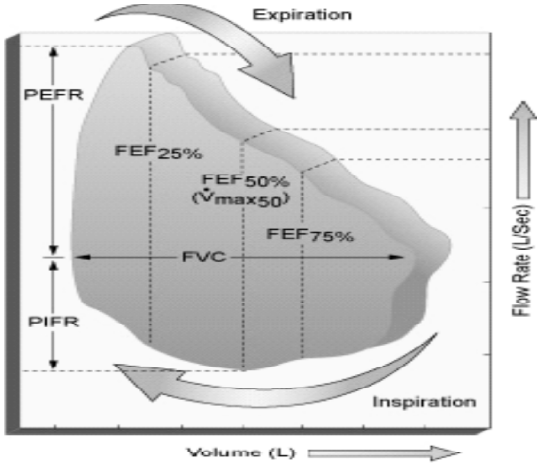

Figure 2.1: Above shows a normal flow-volume loop

A number of dimensions may be obtained from the flow-volume loop, including:

- PEFR- peak expiratory flow rate.
- PIFR -Peak breathe rate of flow.
- FVC - Forced vital capacity.
- FEVT - Forced Expiratory Volume-Time
- FEV1- Volume that has been exhaled at the end of the first second of forced expiration
- FEV1/FVC magnitude relation
- FEF- Forced expiratory flow (at 3 levels 25%, 50%, and 75 %)

Knowledge discovery methods on medical datasets

Preprocessing before data mining:

- Cleaning, Integration and Selection of data

Data mining and knowledge gathering:

- Data mining once the preprocessing.
- Pattern identification and
- Knowledge presentation.

Data mining techniques square measure extensively won't to motivate category description, association, classification, clustering, prediction and statistic analysis.

Data mining in medical field: data processing provides automatic pattern detection for substantial quantity of unsure medical knowledge and tries to disclose patterns in knowledge that square measure powerful to spot with standard applied mathematics ways. The appropriate data processing techniques may be analyzed and applied in line with the relevant domain demand. Few Classification data processing techniques are given below.

Associative rule mining: Association rules square measure shaped by analyzing knowledge for frequent if/then patterns and victimization the factors support and confidence to acknowledge the foremost necessary relationships. Associate degree association rule has 2 sections, associate degree antecedent (if) and a resulting (then). Associate degree antecedent is a component found within the knowledge. Resulting is a component that's earned together with the antecedent. Associate degree correct prediction of the worth of a goal side can make a comeback decision-making method. IF-THEN rules square measure expeditiously utilized in several of the info mining approaches. This provides distinction among the discovered data at a high level of abstraction in Figure 2.2.

Decision tree: A choice tree may be a structure that features a root node, branches, and leaf nodes. Every internal node represents a check on associate degree attribute, every branch denotes the end result of a check and every leaf node represents a category label. The top node within the tree is that the root node. Set of conditions structured hierarchically in such some way that the last word call may be determined by following the conditions that square measure glad from the basis of the tree to its individual leaves. They simply understood and enforced by the researchers all told disciplines. The choice tree model use the conditional rules for classification that square measure simple to know by the user. call tree is one in every of the foremost used data processing techniques, that beat up one table, and over one attribute at a time the choice tree in **Figure 2.3**. Three is constructed supported the coaching knowledge set in **Table 2.1.** Each row within the table corresponds to a patient record, which may be a knowledge instance. The info set contains 3 predictor attributes, comparable to

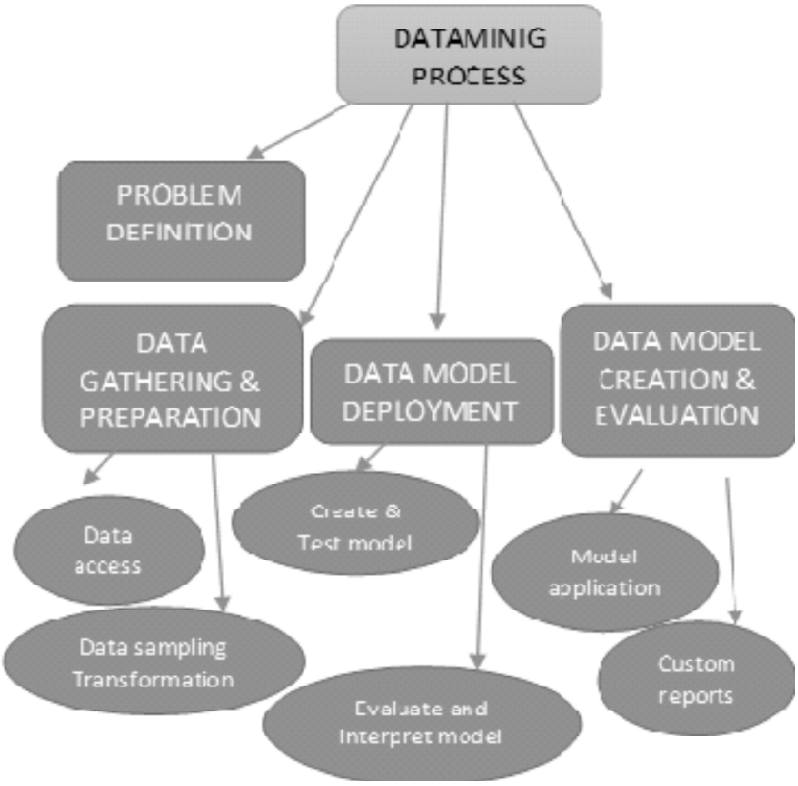

Figure 2.2: Process of Data Mining

Age, Gender, and Symptom intensity. The resultant attribute within the table is sickness, that establish whether or not the associated patient have sickness or not. The resultant attribute values square measure foreseen in line with the values of symptom intensity. The choice tree so classifies the on top of knowledge set within the Table 2.1.

Table 2.1
Data set used to build decision tree of Figure 2.3

Age	*Gender*	*Symptom Intensity*	*Disease (Resultant)*
31	Male	medium	Yes
32	Male	high	Yes
30	Female	medium	Yes
43	Female	high	Yes
30	Female	low	No
22	Male	low	No
19	Female	low	No
28	Male	medium	No

3. CONCLUSION

With this paper, numerous strategies and methodologies utilized in spirometry knowledge classification were analyzed. Numerous parameters like PEFR, PIFR, FEV1, FVC, FEC, MVV and FEV1/FVC magnitude relation were used for the classification of knowledge 2 curves gift within the spirometry tests that square measure flow-

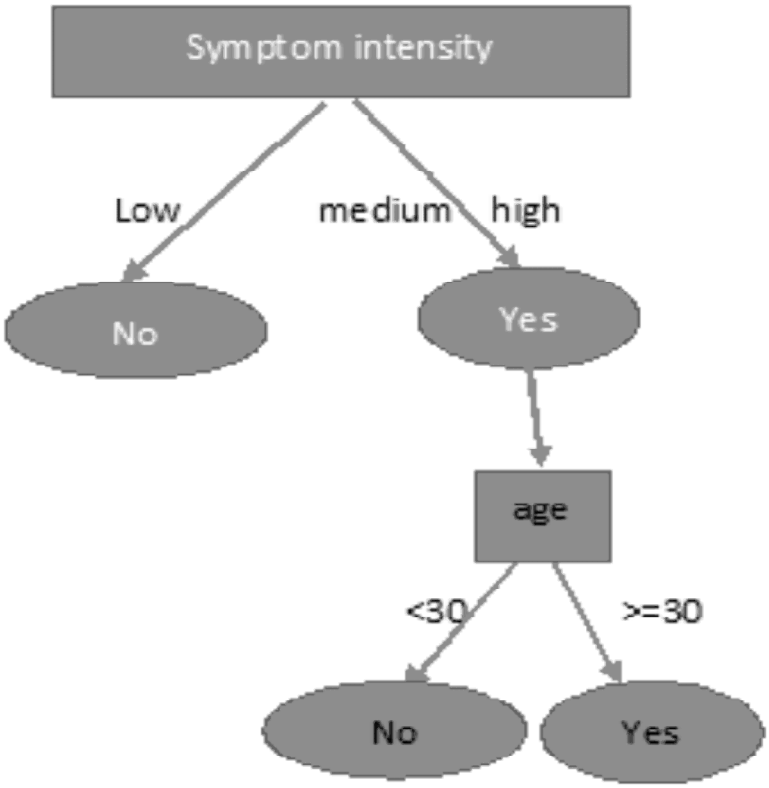

Figure 2.3: Showing the decision tree

volume loop and volume-time curve square measure wide analyzed victimization neural networks. As shown on top of connected work a large sort of researchers have used Multilayer Perception Neural Network (MLPNN) and repeated neural network for detection and classification of the complete 3 patterns of metabolism abnormalities. The feed forward neural networks get flow volume curves. The activation operates representing the pressure – volume relationship of the respiratory organ were used to get the pressure and resistance parameters. A number of the researchers used ANOVA technique and Spline operate. They even have used Support Vector and Regression algorithmic program to classify Spirometry knowledge and predict severity of respiratory organ sickness. Few of the researchers have used call tree and associative rule for spirometry knowledge classification. This paper analyses each neural network and data processing technologies for spirometry knowledge analysis, detection and classification.

4. FUTURE SCOPE

In Spirometry knowledge Classification multilayer perception neural networks (MLPNN), Regression algorithmic program ANOVA technique square measure widely used .The feed forward neural networks square measure wont to generate flow – volume curves. In brief each neural network and data processing technologies may be used for analysis, detection and classification of spirometry knowledge. In future alternative data processing techniques like associative rule mining, call tree and K-Means clump algorithmic program may be wide used for additional totally different and correct result.

REFERENCES

[1] Tsai M J, Pimmel R L, Donoue J F, "Automatic classification of spirometric data", IEEE transaction on biomedical engineering, 1979.

[2] Elman J. "Finding Structure In Time", Cognitive Since. 1990, 14, pp. 179-211.

[3] Abboud S, "Maximum expiratory flow volume curve: mathematical model and experimental results". Med. Eng. Phys. 1995, 17(5), pp. 332-36.

[4] Gaetano P, Marieann H, "Assessment of respiratory mechanism by artificial neural network: an exploratory study". J. Appl. Physiol. 2001, 90(5), pp. 1817-24.

[5] E-handbook of breeze-suite *http://www.medgraphics.com* 2005.

[6] Mahesha V, Ramakrishna S,"Assessment and classification of normal and restrictive respiratory conditions through pulmonary function test and neural network "Journal of Medical Engineering & Technology. 2007, 31(4), pp. 300-04.

[7] Baemani M H, "Detection of Respiratory Abnormalities Using Artificial Neural Networks", Journal of Computer Science. 2008, 4(8), pp. 663-67.

[8] Jafari S, Arabalibeik H, Agin." Classification of normal and abnormal respiration patterns using flow volume curve and neural network" Health Informatics and Bioinformatics (HIBIT), 5th International Symposium,2010Apr, pp. 110-13.

[9] Kavitha A, "Prediction of spirometric Forced Expiratory Volume data using Support Vector Regression", MEASUREMENT SCIENCE REVIEW, 2010, 10(2), pp. 63-67.

[10] Anadan K, "Evaluation of Flow- Volume Spirometric test Neural Network Based prediction and Principal Component Analysis" J. Med SYST. 2011, 35(1), pp. 127-33.

[11] Zhanqi Zhao, Knut Möller Barbara Vogt, Inéz Frerichs "Customized evaluation software for clinical trials: an example on pulmonary function test with electrical impedance tomography "978-1-4673-2971-2/13/$31.00 ©2013 IEEE.

[12] Roberto Maestri and Gian Domenico Pinna "A computer Application for the Investigation of Cardiac Autonomic Effects of Sleep-disordered Breathing in Heart Failure Patients" 8th Conference of the European Study Group on Cardiovascular Oscillations (ESGCO 2014).

[13] Chathuri Daluwatte' , Christopher G. Scully', George C. Kramer2, David G. Strauss' "A Robust Detection Algorithm to Identify Breathing Peaks in Respiration Signals from Spontaneously Breathing Subjects" ISSN 2325-8861, 2015.

[14] Gyu-Yil Choi, Sang-Sub Park, "An Analysis of Actual Conditions with the Infectious Disease of Middle East Respiratory Syndrome Coronavirus (MERS-CoV) Spreading in South Korea", Indian Journal of Science and Technology, 2016 July, 9(25), pp. 1-7.