



Blood Oxygen Level Screening using Data Mining

Reshma Malik
Assistant Professor

Department of Information Technology
Thadomal Shahani Engineering College
Mumbai, India

Jayant Gadge

Associate Professor

Department of Computer Engineering
Thadomal Shahani Engineering College
Mumbai, India

ABSTRACT

Sleep has to serve an important function to restore and recreate physical and mental performance. Current lifestyle attracts more attention towards sleep disorder. Hence sleep disorder is the subject of interest. This is now a common diagnostic process for health worldwide.

The Polysomnography analyzer uses data mining technique to analyze patient blood oxygen level. Chronic Obstructive Pulmonary Disease (COPD) sleep disorder is caused by reduction of oxygen level in blood. The Polysomnogram recording is an 6 to 8 hrs sleep study. 100 polysomnography recordings of 6 to 8hrs each are used to analyze the results. Work emphasis on use of association rule mining to extract interesting patterns from the Polysomnogram recording.

General Terms

Association rule mining.

Keywords

Association rule mining, COPD, Data mining, Polysomnography, Sleep disorder.

1. INTRODUCTION

Sleep is a physical as well as mental resting state in which an individual become comparatively inactive and not aware of the environment. That means sleep is a partial detachment from the world where most of the external stimuli are senseless [1]. Hence, Sleep is a necessary and extremely important biological function.

In many cases, individuals with good health tend to sleep well, whereas those with problems of physical and/or mental health show sleep related disorders. Even minimum sleep loss can take a toll on individuals mood, energy, efficiency and ability to handle stress. Ignoring sleep disorders can lead to poor health, accidents, poor job performance, and relationship stress. In order to feel best, stay healthy, and give optimal performance, sleep is necessity it is not a luxury.

Polysomnography, also known as a sleep study, is a multi parametric study of sleep and used as a diagnostic tool to identify sleep disorder. The test result is called a polysomnogram abbreviated as PSG. PSG measures different physiological parameters while an individual is asleep. During 1970s PSG was used first time to detect sleep irregularities at Stanford University. PSG comprises signals from neurological and repertory sensors. The neurological signals are electroencephalogram (EEG), electromyogram (EMG), and electrooculogram (EOG) The PSG present detailed time-series

representation of sleep [2]. The polysomnogram signals are stored on epoch-by-epoch basis (epoch is 30 sec duration). Polysomnography recording is on average, 6–8 hr measurements. Rechtschaffen and Kales model is use to classified sleep stages as: wake, 1, 2, 3, 4, and rapid eye movement (REM) [3], [4]. The various respiratory channels used in PSG includes Thorax, Abdominal, Snore, Spo₂, Pulse, Flow, Position etc.

The design of Polysomnogram Analyzer is given as below

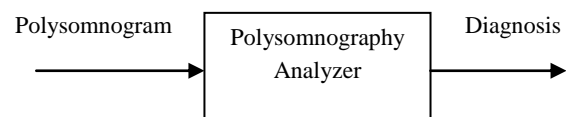


Fig.1 Block Diagram of 'Polysomnography Analyzer'

In this paper, Literature review and all the related work related to COPD and Oxygen Desaturation is discussed .In section 2. In section 3, design methodology is discussed .In section 4, experimental results and graphs are discussed whereas in section 5, conclusion is presented.

2. LITERATURE REVIEW

COPD is a progressive disease that causes difficulty in breathing. Most of the time it gets worse over the period time. COPD can cause coughing, wheezing, shortness of breath, chest tightness and other symptoms. Cigarette smoking is the one of the leading reason for COPD. Other reasons include air pollution, chemical fumes, or dust [5].

The air that individual breathe goes down into tubes in lungs called bronchial tubes or airways. Within the lungs, bronchial tubes branch into thousands of smaller, thinner tubes called bronchioles. These tubes end in bunches of tiny round air sacs called alveoli. Small blood vessels called capillaries run through the walls of the air sacs. When air reaches the air sacs, oxygen passes through the air sac walls into the blood in the capillaries. At the same time, carbon dioxide moves from the capillaries into the air sacs. This process is called gas exchange [5].

The airways and air sacs are elastic. When individual breathe in, each air sac fills up with air like a small balloon. When individual breathe out, the air sacs deflate and the air goes out.



There are various reasons for less air flows in and out of the airway such as elastic quality of the airways and air sacs are lost or walls between the air sacs are destroyed etc.

Table 1 Related Disease with Parameter

Parameter Name	Normal Range	Abnormal Range	Related Disease	Related Channel
SpO ₂	90%-98%	< 90%	COPD	SpO ₂
Oxygen Desaturation Index	< 5 per hour	5-15 mild 15-30 moderate >30 severe	Breathing Problem	SpO ₂

As PSG is comprehensive recording of various sleep related channel. The channel normally referred to identify the COPD is the SpO₂. Table above explains the normal and abnormal ranges of oxygen.

The Oxygen Desaturation Index (ODI) measures the drop age of oxygen in blood by 4 or more percentage. The drops in oxygen levels called Desaturation. These drops are associated with collapses of the upper airway, this events called apnea or hypopnea. The Obstructive Sleep Apnea (OSA) normally consider as leading cause for Oxygen Desaturation.

3. DESIGN METHODOLOGY

The data used in analysis is collected from different patient using a Polysomnogram instrument called as sleep lab machine. The recording is stored in European Data Format (EDF). The preprocessing block is used to convert the edf data to CSV format [6].

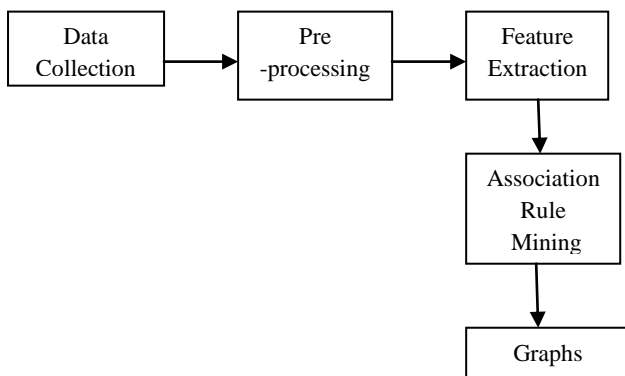


Fig. 2 Detail Block Diagram of 'Polysomnography Analyzer'

The normal and abnormal range is extracted by using feature extraction block. To perform exploratory data analysis association rule mining is the best technique in data mining. The idea behind the use of association rule mining is to state presence of certain attribute value condition for presence of certain other attribute condition [7]. The dataset used for analysis comprises of objective information collected from various healthy and unhealthy individuals. Polysomnography recordings for 100 patients were analyzed which is collected from Pulmonix Health Care and Respitec. Polysomnography leads to the collection of large amounts of data that needs to be analyzed for diagnosis.

The basic assumption for designing polysomnography analyzer is all sleep lab machine generates 'edf' file.

3.1 The step used in Analysis Process

1. Select a patient record.
2. If a patient ASCII CSV is not present regenerate it.
3. Load CSV to matlab processor.
4. Apply Rule base to the data.
5. Generate graphs.

3.2 The flowchart used in Analysis

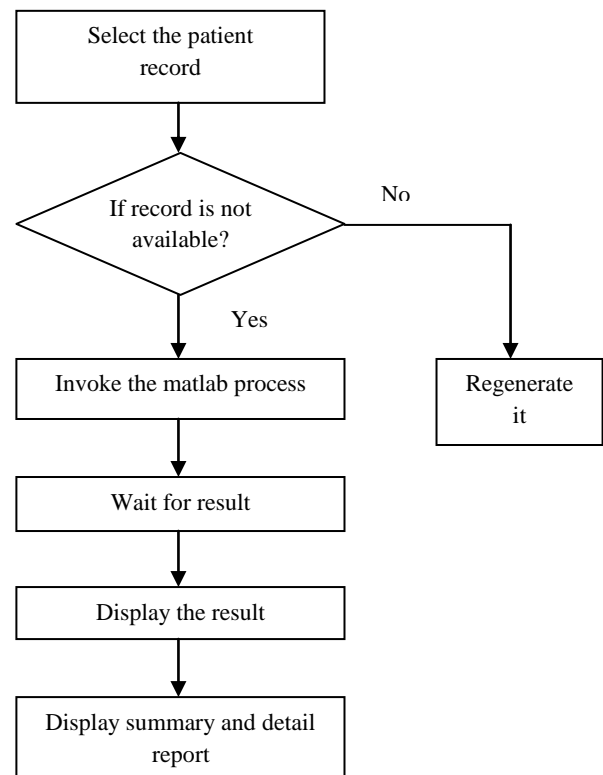


Fig.3 Flowchart used in analysis

The polysomnogram recording contains the high frequency as well as the low frequency channels. The entire polysomnography recording is stored according to the channels to reduce the time required to read the file. The time requires for reading the data file depends on the sampling rate and reduced to great extent. The table listed below represents the channel with its sampling rate.



Table 3 channel and sampling frequency

Channel Name	Sampling Rate
Spo2	3Hz

The conversion of polysomnogram recording to ASCII CSV Format is done using a command line utility. Four ASCII output files are created in the same folder as the EDF file [6]. The various files are:

1. filename_header.txt contains the general header of the EDF file.
2. filename_signals.txt contains all signal headers of the EDF file.
3. filename_data.txt contains all signals.
4. filename_annotations.txt contains all annotations.

The filename_data.txt contains a separate line for Sampletime. Each line contains the comma-separated values of the Sampletime and of all Samples that were taken at that time. If the value is not present then it is left empty, but the comma-separator is present. Sampletime is expressed in seconds, while sample values are expressed in the Physical dimensions that are specified in filename_signals.txt. All values have a dot (.) as decimal separator.

4. RESULTS

1. The Application Interface allows user to select the EDF file through the common dialog box. The selected file is processed to convert the ASCII CSV file. The four CSV files are created and they are named as traces_header.txt, traces_annotation.txt, traces_signals.txt and traces_data.txt.
2. The split function separates each and every value of ASCII CSV view file. The data cleaning and data validation is done through the pre-process function.
3. The traces_header .txt gives the information about total number of channels used for recording the polysomnogram recording. One of the signals is spo₂. The signal name is extracted from traces_signals.txt file whereas the signal data is available in traces_data.txt.
4. The extracted spo₂ signal is further processed by an application interface according to the spo₂ range defined in table 1. The further processing needs file to be converted to an .ariff file which is the file format of WEKA. Hence the generated an .ariff file processed by WEKA tool.
5. The default algorithm used in association rule mining is Apriori algorithm. The support and confidence are used as parameters for generating rule base. The default setting for rule generation in WEKA is 10 best rules. It begins with a minimum support of 100% and decreases this in step of 5%. It stops when it generates 10 rules or support falls below 10% or
6. minimum confidence is reached to 90%. A low support rule is likely to be uninteresting and occur by chance.

The graph represents SpO₂ percentage with respect to time. The oxygen Desaturation level for an entire 8hr study is marked on the graph as given below.

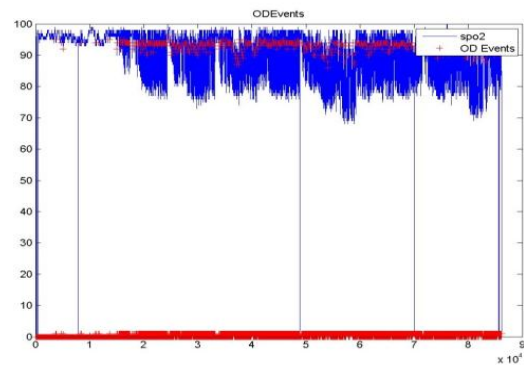


Fig.4 Graph explaining Oxygen desaturation.

5. CONCLUSION

The COPD disease is life threatening disease. It has been observed that in COPD patient, the oxygen level is low and carbon dioxide level is high. COPD reduces the amount of oxygen reaching to lungs and transferred to the blood, while anemia which is a common problem associated with COPD patients limits the amount of oxygen that the blood can carry as a result anemia and COPD decrease exercise capacity in patient. COPD diagnosis is often missed or delayed due to lack of awareness.

The early and regular detection of oxygen and CO₂ level in body is essential to reduce its impact. The polysomnography analyzer can be used in diagnosis centre for initial analysis. It helps to detect oxygen level in the body Hence, Polysomnography analyzer can be useful in the area of healthcare.

6. REFERENCES

- [1] Sleep introduction, <http://www.aasmnet.org>
- [2] N. Butkov, *Atlas of Clinical Polysomnography*, vol. 2, Medford, Synapse Media, 1996.
- [3] A. Rechtschaffen and A. Kales, *A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects*, U.S.Dept. Health, Education, and Welfare, NIH, Publ. 204, Washington, DC,1968.
- [4] M. A. Carskadon and A. Rechtschaffen, "Monitoring and staging human sleep," in M. H. Kryger, T. Roth and W. C. Dement, *Principles and Practice of Sleep Medicine*. 3rd, ed. Philadelphia,PA: Saunders,2000.
- [5] COPD introduction <http://www.nhlbi.nih.gov/health/health-topics/topics/copd/> access on May 2013.
- [6] EDF/BDF converter <http://www.teniz.net/edf2ascii/access> on Dec2012.
- [7] E. Frank and I. H. Witten, *Data Mining: Practical Machine Learning Tools and Techniques*, 2nd ed., San Mateo, CA: Morgan Kaufmann,2005.
- [8] WEKA Manual for version3-7-9,University of Waikato,Hamilton,New Zealand,February19,2013.
- [9] Embla titanium Clinical Manual, Thornton, USA,2011.