Research Paper

ANALYSIS OF BREATHING PROBLEM ON SMOKERS AND NONSMOKERS USING MACHINE LEARNING

P.J. MAHIBA Assistant Professor, Department of computer Application, St. John's college of Arts and science, Ammandivilai

DR. R. KAVITHAJABA MALAR, Department of computer Science St. John's college of Arts and science, Ammandivilai

ABSTRACT: Common causes of feeling short of breath are: lung problems, such as asthma and chronic obstructive pulmonary disease (COPD) heart problems, such as a cardiovascular disease and heart failure. Infections in the airways, such as croup, bronchitis, pneumonia, COVID-19, the flu and even a cold.The problem of screening a lung is a major issue in the medical area. One of the main goals of this research is to analyze the breathing disorder on smokers and non smokers using decision tree algorithm. Developing an automatic solution to find the abnormality in lung tissues and to predict about which kind of disorder has been occurred. The aim of this research is to analyze and identify the type of breathing problem in both smokers and non smokers.

KEYWORDS: Algorithm, Cancer, Decision Tree, Lung.

I. INTRODUCTION

Breathing problems are when you feel you can't get enough air, your chest feels very tight, you are breathless or you feel like you're being suffocated. You might feel short of breath if you are obese or if you have just done some strenuous exercise. Breathing is the physical process of inhaling oxygen and exhaling carbon dioxide. The mechanism of breathing involves two main processes: inspiration and expiration. Inspiration occurs when the diaphragm and the external intercostal muscles contract.. Many breathing problems are long-term (chronic). These include chronic sinusitis, allergies, and asthma. They can cause symptoms such as nasal congestion, a runny nose, itchy or watery eyes, chest congestion, coughing, wheezing, trouble breathing, and shallow breathing.Breathing uses chemical and mechanical processes to bring oxygen to every cell of the body and to get rid of carbon dioxide. Our body needs oxygen to obtain energy to fuel all our living processes. Carbon dioxide is a waste product of that process. Stress and strong emotions can cause breathing difficulties such as shortness of breath and rapid breathing. In a healthy person, the effects caused by stress on the lungs are usually not dangerous, but for people with breathing problems, it could worsen the symptoms. Various factors affect a person's breathing rate, including injuries, exercise, emotion, mood, and a range of medical conditions. Smoking can cause lung disease by damaging your airways and the small air sacs (alveoli) found in your lungs. Lung diseases caused by smoking include COPD, which includes emphysema and chronic bronchitis. Cigarette smoking causes most cases of lung cancer.

II.PROPOSED METHOD

Data mining is the process of automatically collecting large volumes of data with finding hidden patterns and analyzing the relationships between numerous types of data to develop predictive

Research Paper

models. In this research, the classification techniques are used to analyze the lung cancers in smokers and non smokers. Classification and prediction are two forms of data analysis that can be used to extract models describing important data classes or to predict future data trends. Decision tree is used for predicting the Lung Cancer Disease from the given data set instances. The input parameters are processed and rules are generated for the analysis.
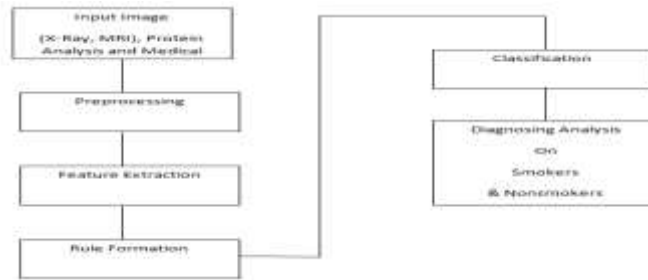

Fig.1: Proposed System

Input x-ray and MRI images, Pro and M-Reco

Let L = {set of frequent itemsets}

Set k = 1, L = φ ; C1 = candidate itemset

L1 = frequentitemset ;Produce 1 (C1) and (L1)

do while (Lp-1 ! = φ)

{

Produce k – itemsets (Ck) from frequent itemsets(Lk-1)

Produce frequent itemsets (Lk) from k – itemsets(Ck)

Increment K

}

Produce frequent itemsets (L)

Apply association rule (R) using associationgeneration formula.
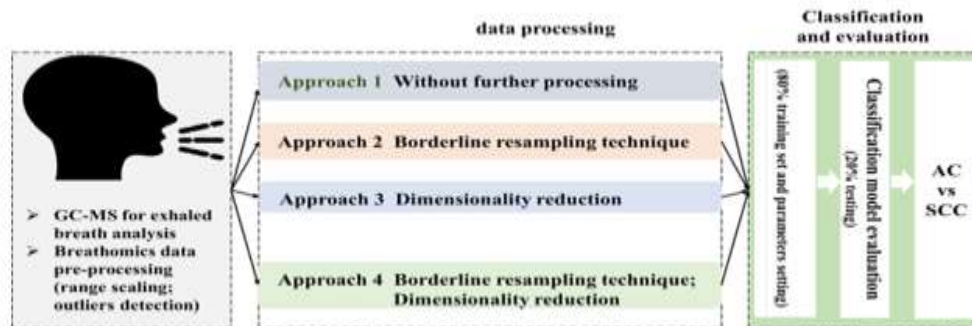

Fig.2: Classification Method

Decision tree is used for classification. The attributes with valid values are extracted. Eleven attributes are identified Gender, age, race, reformed smokers, length of smoking, pack-year,

Research Paper

Prot_stat, xray-imp, mri-len,mri-wid, mri-dep. Interesting rules have been generated. Apriori Algorithm is used for association mining. There are lots of evidences proving that smokers with a particular mutation have a dramatically higher risk to develop lung cancer. Therefore the comparison of the smokers' genes expression with the non-smokers' could be helpful in order to discover the effect of smoking on airway gene expression. In this study we used microarray data of large epithelium lung cells to compare normal non-smokers with normal smokers.

III.CONCLUSION

Data mining plays a major role in extracting the hidden information in the medical data base. The data preprocessingis used in order to improve the quality of the data. This model is built based as a test case on the UCI repository dataset. The experiment has been successfully performed with several data mining classification techniques and it is found that the Naive Bayes algorithm gives a better performance over the supplied data set with the accuracy of 83.4%. It is believed that the data mining can significantly help in the Lung Cancer research and ultimately improve the quality of health care of Lung Cancer patients. It can also be implemented using several classification Techniques.

REFERENCE

[1] Jiawei Han and MichelineKamber, "Data Mining Concepts and Techniques", Second Edition, University of Illinois at UrbanaChampaign, 2006.

[2] Lawrence A. Loeb, Virginia L. Ernster, Kenneth E. Warner, John Abbotts, and John Laszlo "Smoking and Lung Cancer", on July 17, 2014.

[3] I. T. T. HIGGINS, "Commentary On "Possible effects Noccupational Lung Cancer From Smoking Related Changes In The Mucus Content Of The Lung", Department of Epidemiology, School of Public Health, The University of Michigan, Ann Arbor, 22 November 1982.

[4] Kawsar Ahmed1, Abdullah-Al-Emran2*, Tasnuba Jesmin1, Roushney Fatima Mukti2, MdZamilur Rahman1, Farzana Ahmed3, "Early Detection of Lung Cancer Risk Using Data Mining", Asian Pacific Journal of Cancer Prevention, Vol 14, 2013.