INTERNATIONAL RESEARCH JOURNAL IN ADVANCED ENGINEERING AND TECHNOLOGY (IRJAET) E - ISSN: 2454-4752 P - ISSN : 2454-4744
VOL 4 ISSUE 2 (2018) PAGES 2819 - 2825
RECEIVED : 25.02.2018 PUBLISHED : 21.03.2018
March 21, 2018

A survey on Health Care analysis to Detect Lung cancer using data mining Techniques

[1]A.V. Sriramvignesh, Department of Computer Application,
Ganadipathy Tulasi's Jain Engineering college, Vellore.

[2]S. Sathish Kumar, Department of Computer Application,
Ganadipaty Tulasi's Jain Engineering college, Vellore.

Abstract

The health care industry has a large amount of medical data. These data are collected from the patient who have undergoes any kind of medical treatment or test. In this field cancer is one of the top most disease which can affect most part of human body . The world statistic report reveals that lung cancer is the second most place of cancer leads to death. Here we use some Descriptive and predictive Data mining techniques to detect early stage of lung cancer using classification, clustering ,association, prediction. objective of our paper is to review the past work done using the data mining techniques in prediction of lung cancer.

Keywords: Lung cancer, classification, clustering.

1. INTRODUCTION

Cancer is becoming one of the dead line disease in world. our body has many types of cells which undergoes mitosis and grows several time in a particular time period maintain normal function. At the time cells affected and create an abnormal cells. The cells which split usually are called cancer. these cells can be split into same part or other organ of the body. cancer is also called carcinoma. Where the early stage of diagnosis of cancer is still challenging to doctor.

Lung growth is a malignancy that begins in the lungs. The event of lung disease has expanded quickly and moved toward becoming the most widely recognized tumor in men in many nations. Smoking is by a long shot the most essential preventable reason for malignancy on the planet. On the off chance that the first lung tumor has spread, a individual may feel side effects in different places in the body. Normal spots for lung malignancy to spread incorporate other parts of the lungs in Tumor, lymph hubs, bones, mind, liver. The frequency of lung growth is firmly connected with cigarette smoking, with around 90% of lung diseases emerging because of tobacco utilize. The danger of lung growth increments with the quantity of cigarettes smoked after some time. A great many people realize that smoking causes malignancy, however may not understand what number of nonsmokers get lung growth, as well. The reasons for this work is finding the hazard factor of lung tumor and order the smokers and non-smokers who are altogether caused by lung tumor by utilizing the information mining Technique. For detect a disease

number of tests should be required from the patient. By using the data mining number of test will be reduced. this reduce play an important role in time and performance.

2. RISK FACTORS

- smoking (cigarette, pipe, and cigar)
- radon (a naturally occurring gas)
- asbestos (natural mineral used in construction),
- some metals (cadmium, chromium, arsenic),
- exhaust from diesel engines, radiations, chemicals in paint and certain diseases that affect the lung (e.g. tuberculosis).
- Family history of cancer can also put a person at greater risk.

3. LITERATURE SURVEY

i. P.Ramachandran and T.Bhuvaneswari [10] they identify the genetic and environmental factors to detect and prevent lung cancer .This research uses data mining technology such as classification, clustering and prediction to identify potential Lung cancer patients. The gathered data is preprocessed, fed into the database and classified to yield significant patterns using decision tree algorithm.

ii. N.Naveenkumar and G.Selvavinayagam[3] Mining Techniques for clinical expert system and predicting and treating lung cancer with big data .This has described about the general facts about the lung cancer using a information mining techniques. The important role of Neural Network and Naïve Bayes algorithms are discussed to prevent the cause of lung cancer.

iii. J.Jamera banu[4] This paper has successfully performed with several data mining classification techniques and they 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.

iv. Neha Panpaliya, Neha Tada[5] This paper they conclude that using the combination of neural network classifier along with binarization and GLCM will increase the accuracy of lung cancer detection process. By using this system will also decrease the cost and time required for cancer detection and also if the patient is not detected with the lung cancer the system will proceed further for the prediction

4. DATA MINING TECHNIQUE

Information mining is the procedure of consequently gathering extensive volumes of information with the target of finding covered up designs and investigating the connections between various kinds of information to create prescient models. The grouping procedures and forecast are two types of information investigation that can be utilized to remove models depicting imperative information classes or to foresee future information patterns. Such examination can assist give us a superior comprehension of the information on the loose.

A.CLUSTERING

Clustering is an information mining method that makes valuable or supportive group of substance that have comparative component utilizing mechanical system. From grouping, bunching system characterizes the classes and keeps questions in them, as in order objects are alloted into predefined classes. For example prediction of lung cancer disease by using clustering obtain cluster or state that list of patients have same risk factor.

B. *ASSOCIATION*

Association is outstanding amongst other information mining procedures. In affiliation, an example is uncovered in view of a relationship of a specific thing on different things of a similar activity. For example, the association technique is used in finding lung cancer disease prediction as it say to us the relationship of dissimilar attributes used for analysis and sort out the patient with all the risk factor which are prediction of disease.

C. PREDICTION

Utilized revenue driven forecast, relationship among reliant and free factors and connection between autonomous factors. For instance, forecast investigation procedure can benefit from outside assistance in deal to anticipate benefit for the future if consider deal is an autonomous variable, benefit could be a reliant variable.

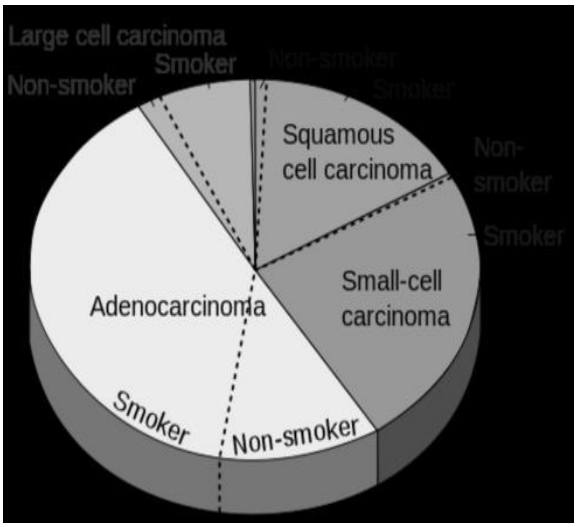
5. DATA MINING PROCESS

In the information digging strategies are for removing designs from information. The examples that can be found rely on the information mining process connected. By and large there are two strategies for information mining undertakings, expressive information mining assignment is that portray the general properties of the current information, and prescient information mining errand is that endeavor to do forecasts in light of accessible information. Information mining should be possible on information which are in quantitative and interactive media. Information mining applications can show distinctive sort of parameters to inspect the information. They incorporate affiliation (designs where one occasion is associated with another occasion), arrangement or way examination, order (distinguishing proof of new examples with predefined targets) and bunching (gathering of indistinguishable or comparable items).

6. CLASSIFICATION PROCESS

Characterization technique is utilized to investigate the Somker's and non smoker's hazard factors in view of every cell of human and the phases of lung disease with the assistance of Weka device. Order of lung growth in view of cell carcinoma.

INTERNATIONAL RESEARCH JOURNAL IN ADVANCED ENGINEERING AND TECHNOLOGY (IRJAET) E - ISSN: 2454-4752 P - ISSN : 2454-4744
VOL 4 ISSUE 2 (2018) PAGES 2819 - 2825
RECEIVED : 25.02.2018 PUBLISHED : 21.03.2018
March 21, 2018



a)*ADENOCARCINOMA*

Adenocarcinoma is a typical histological type of lung disease. Almost 40% of lung tumors are adenocarcinoma, which typically starts in fringe lung tissue. Most common from lung cancer.

b)SQUAMOUS CELL CARCINOMA

These diseases begin in early forms of squamous cells, which are level cells that line within the aviation routes in thelungs. They are frequently connected to a background marked by smoking and have a tendency to be found amidst the lungs, close to a bronchus.

c)LARGE CELL CARCINOMA

This kind of growth represents around 10% to 15% of lung tumors. It has a tendency to develop and spread rapidly, which can make it harder to treat. A subtype of substantial cell carcinoma, known as substantial cell neuroendocrine carcinoma, is a quick developing disease that is fundamentally the same as little cell lung tumor.

d) *SMALL CELL CARCINOMA*

Little cell carcinoma frequently begins in the bronchi close to the focus of the chest, and it tends to spread broadly through the body genuinely right on time over the span of the sickness.

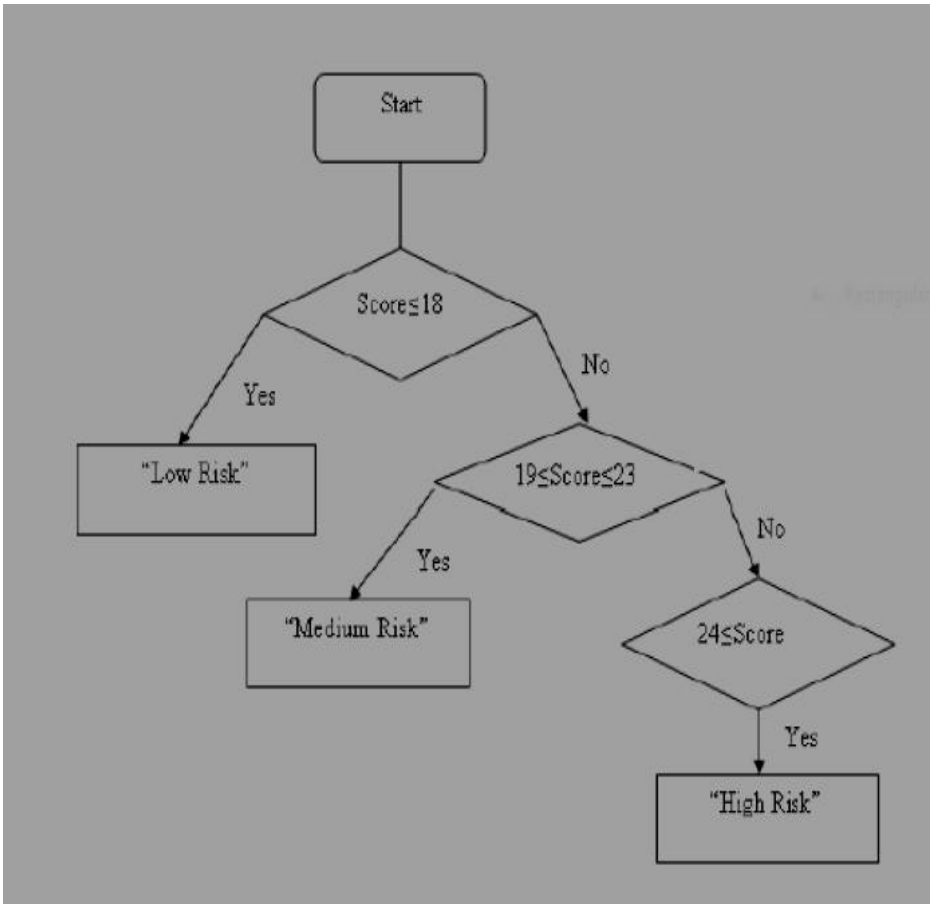
7. PREDICTION PROCESS

Cancer research is generally clinical and/or biological in nature, data driven statistical research has become a common complement. Predicting the outcome of a disease is one of the most interesting and challenging tasks where to develop data

INTERNATIONAL RESEARCH JOURNAL IN ADVANCED ENGINEERING AND TECHNOLOGY (IRJAET) E - ISSN: 2454-4752 P - ISSN : 2454-4744
VOL 4 ISSUE 2 (2018) PAGES 2819 - 2825
RECEIVED : 25.02.2018 PUBLISHED : 21.03.2018
March 21, 2018

mining applications. As the use of computers powered with automated tools, large volumes of medical data are being collected and made available to the medical research groups. As a result, Knowledge Discovery in Databases(KDD), which includes data mining techniques, has become
a popular research tool for medical researchers to identify and exploit patterns and relationships among large number of variables, and made them able to predict the outcome of a disease using the historical cases stored within datasets.

There is no solution for malignancy after totally influenced. Demise is unavoidable. So the capacity to foresee Lung growth assumes an essential part in the analysis procedure. In this paper we have proposed a successful Lung malignancy expectation in light of information mining. This lung growth hazard forecast framework ought to demonstrate accommodating in discovery of a man's inclination for lung disease. We will consider different hazard factors which incorporates age, sexual orientation, innate, past wellbeing examination, utilization of against extremely touchy medications, smoking, sustenance propensity, physical movement, heftiness, tobacco, hereditary Risk, condition, mental injury, take-up of red meat, adjust abstain from food, hypertension, coronary illness, over the top liquor, radiation treatment and incessant lung ailments. Different calculations will be utilized, for example, Figure underneath outlines the working of frequent item sets that occur throughout the data base and have a identify link to lung cancer status are mined. The data is fed into the decision tree to obtain That patient is related to lung cancer and non cancer data sets.



CONCLUSION

The objective of our work is to provide a study of different data mining techniques that can be employed in automated disease prediction systems. Using descriptive and predictive techniques of data mining and it is defined in this work which has emerged in recent years for efficient and effective disease. The analysis shows that different technologies are used in all the papers with taking different number of attributes. Many patients did not know Lung cancer disease in early stage, because of the lack of awareness. The emphasis of this work is to find the target group of people who needs further Screening for Lung cancer disease, So, different techniques used shown the different accuracy depends upon number of attributes taken and tool used for implementation. Although applying data mining techniques to help health care professionals in the detecting of Lung Cancer disease .

REFERENCE

[1] Mohammad Taha Khan, Dr. Shamimul Qamar and Laurent F. Massin, A Prototype of Cancer/ Disease Prediction Model Using Data Mining, International Journal of Applied Engineering Research, 2012.

[2] J.Jamera banu, Study of Classification Algorithm for Lung Cancer Prediction, International Journal of Innovative Science, Engineering & Technology, Vol. 3 Issue 2, February 2016.

[3] Naveenkumar.N and G. Selvavinayagam,2015.Mining Techniques for clinical Expert System and Prediction and Treating Lung cancer with Big data,ISSN:2347-8586,

[4] J.Jamera banu, A Study on Mining Lung Cancer Data for Increasing or Decreasing Disease PredictionValue by Using Ant Colony Optimization Techniques Special Issue Published in Int. Jnl. Of Advanced Networking and Applications (IJANA) Page 150.

[5] Neha Panpaliya, Neha Tadas, Surabhi Bobade, Rewti Aglawe, Akshay Gudadhe, A Survey On Early Detection And Prediction Of Lung Cancer, International Journal of Computer Science and Mobile Computing, Vol.4 Issue.1, January- 2015, pg. 175-184.

[6] Mr. P. Thangaraju, R. Mehala, Novel Classification based approaches over Cancer Diseases, International Journal of Advanced Research in Computer and Communication Engineering Vol. 4, Issue 3, March 2015.

[7] H.Bharathi, TS. Arulananth, A Review of Lung cancer Prediction System using Data Mining Techniques and Self Organizing Map (SOM), International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 10 (2017) pp. 2190-2195 © Research India Publications. http://www.ripublication.com

[8] Abhishek taneja, Disease Prediction System Using Data Mining Techniques, Oriental Scientific Publishing Co., India, 2013.

[9] M. Anbarasi, E. Anupriya, N.ch.s.n.Iyengar, Enhanced Prediction of Disease with Feature Subset Selection using Genetic Algorithm, International Journal of Engineering Science and Technology,2010.

[10] T.Bhuvaneswari and T.Bhuvaneswari, Early Detection and Prevention of Cancer using Data Mining Techniques, *International Journal of Computer Applications (0975 – 8887) Volume 97– No.13, July 2014*