ISSN (Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology
An ISO 3297: 2007 Certified Organization Volume 8, Special Issue 1, March 2019
7th National Conference on Frontiers in Communication and Signal Processing Systems (NCFCSPS '19)
19th-20th March 2019
Organized by
Department of ECE, Adhiyamaan College of Engineering, Hosur, Tamilnadu, India

Quantitative Analysis Lung Cancer Risk Factors Using Data Mining

Sanku Rajendra Kumar, Dr. Anil Kumar

Research Scholar, Department of Computer Science and Engineering, Sri Satya Sai University of Technology & Medical Sciences, Bhopal, Sehore (M.P.), India

Associate Professor, Department of Computer Science and Engineering, Sri Satya Sai University of Technology & Medical Sciences, Bhopal, Sehore (M.P.), India

ABSTRACT: Lung cancer is one of the most dangerous malignant tumors, with the fastest-growing morbidity and mortality, majority of death caused by uncontrolled growth of cells in any of the tissues or parts of the body. Lung cancer is a deadly disease, but there is a big chance for the patient to be cured if he or she is correctly diagnosed in early stage of his or her case. With a rapid growth of the elderly population in recent years, lung cancer prevention and control are increasingly of fundamental importance, but are complicated by the fact that the pathogenesis of lung cancer is a complex process involving a variety of risk factors. It is hypothetical that cancer is induced by external factors like tobacco, radioactivity, chemical substance or contagious organisms and internal factors like inherited mutations, hormones, resistant situation and mutation that happen from the metabolic process.

Objective: study aimed at identifying risk factors of lung cancer incidence in the elderly and quantitatively analyzing these risk factors' degree of influence using dataming techniques, we integrated multidisciplinary risk factors, including behavioral risk factors, disease history factors, environmental factors, and demographic factors, and then preprocessed these integrated data. We trained deep neural network models in a stratified elderly population. We then extracted risk factors of lung cancer in the elderly and conducted quantitative analyses of the degree of influence using the deep neural network models.

KEYWORDS: Data mining, Lung Cancer, Classification, KNN, SVM, ANN and Risk factor

I. INTRODUCTION

Early detection of cancer to a great extent increases the chances for successful treatment lung cancer is the widespread cancer in both men and Women. Two types of lung cancers are NSCLC and SCLC, the malignant tumors develops when cells in the lung tissue divide and grow without the normal controls on cell death and cell division. Lung cancer treatment decision begins with the stage, or progression of the disease. Using the results diagnostic tests, Lung cancer pathogenesis is a complex process involving various risk factors. Factors such as tobacco consumption, poor diet, occupational exposures and air pollution [5], and drinking water that has a high level of arsenic can increase the risk of occurrence of lung cancer due to tobacco smoking in shaping the descriptive epidemiology of lung cancer. We will review your pathology to confirm you have received the correct diagnosis and staging information and develop a personalized treatment plan and perform comprehensive testing and identify a treatment approach as follows

- **Staging small cell lung cancer** – SCLC make up less than 20 percent of lung cancers and is typically caused by tobacco smoking. It often starts in the bronchi then quickly grows and spreads to other part of the body like lymph nodes. It is stages are classified in two ways
  - Limited stages – The cancer is found in one lung, sometimes including near by lymph nodes.
  - Extensive stages – The cancer has spread to the other lung, the fluid around the lung(the pleura) or other in the body
- **Staging non-small cell lung cancer** – SCLC make up less than 20 percent of lung cancers and is typically caused by tobacco smoking. It often starts in the bronchi then quickly grows and spreads to other part of the

ISSN (Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology
An ISO 3297: 2007 Certified Organization Volume 8, Special Issue 1, March 2019
7th National Conference on Frontiers in Communication and Signal Processing Systems (NCFCSPS '19)
19th-20th March 2019
Organized by
Department of ECE, Adhiyamaan College of Engineering, Hosur, Tamilnadu, India

It is stages are classified in two ways

- **T (tumor)** – This describes the size of the original tumor.
- **N (node)** – This indicates whether cancer is present in the lymph nodes.
- **M (metastasis)** – This refer to whether cancer has spread to other parts of the body, usually the liver, bones or brain.
- **Occult** – Cancer cell are found in sputum, but no tumor can be found in the lung by imaging tests or bronchoscopy or the tumor is too small.

1. **Stage 0:** In this stage is also known as carcinoma in situ. The cancer is tiny in size and has not spread into deeper lung tissues or outside the lungs
2. **Stage 1:** In this stage disease may be present in the underlying lung tissues, but the lymph nodes near your lungs and remain unaffected
3. **Stage 2:** In this stage disease spread further into your lymph nodes near to lungs.
4. **Stage 3:** In this stage cancer is continuing to spread from the lungs to the lymph nodes or nearby structure and organs, such as the heart , trachea and esophagus.
5. **Stage 4:** In this stage, the cancer has metastasized or spread widely around body. It may have spread to your brain, bones or liver.

Management of lung nodules identified by annual low-dose CT screening is generally decided by the most worrying interval change between two previous scans, which could be changes in the diameter or volume of solid lung nodules, an increase in diameter or density of a sub-solid nodule, or the size of a new lung nodule. Trends in lung cancer mortality can be interpreted in terms of different patterns of smoking prevalence in subsequent cohorts of people in various countries. An increase in tobacco consumption was paralleled a few decades later by an increase in the incidence of lung cancer, and a decrease in consumption is followed by a decrease in incidence. Similarly, the temporal lag in trends in female and male lung cancer mortality reflects historical differences in cigarette smoking between subsequent female and male. Luna and colleagues used random forest as an accurate machine learning method to identify known and new predictors of symptomatic radiation pneumonitis, which is a radiotherapy dose-limiting toxicity for locally Fuzzy clustering method to predict lung cancer through continuous monitoring using a new internet of things and to improve health care by providing medical instructions. This study aimed at identifying key risk factors of lung cancer and we focused on multidisciplinary risk factors, such as research problems were how to find the leading causative factors of lung cancer incidence from complex related risk factors and to quantitatively analyze their degree of influence. Our results could help physicians in smoking habit, disease history, radiation exposure, behavioral risk, environmental risk, and medical demographics. Our main preventing lung cancer and taking effective measures for early detection.

II. RESEARCH METHODOLOGY

The research methodology proposed in order to develop an intelligent prediction model for prediction of severity of breast cancer from historical data. The model methodology is embedded within the overall research methodology representation as diagram.

ISSN (Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology
An ISO 3297: 2007 Certified Organization Volume 8, Special Issue 1, March 2019
7th National Conference on Frontiers in Communication and Signal Processing Systems (NCFCSPS '19)
19th-20th March 2019
Organized by
Department of ECE, Adhiyamaan College of Engineering, Hosur, Tamilnadu, India

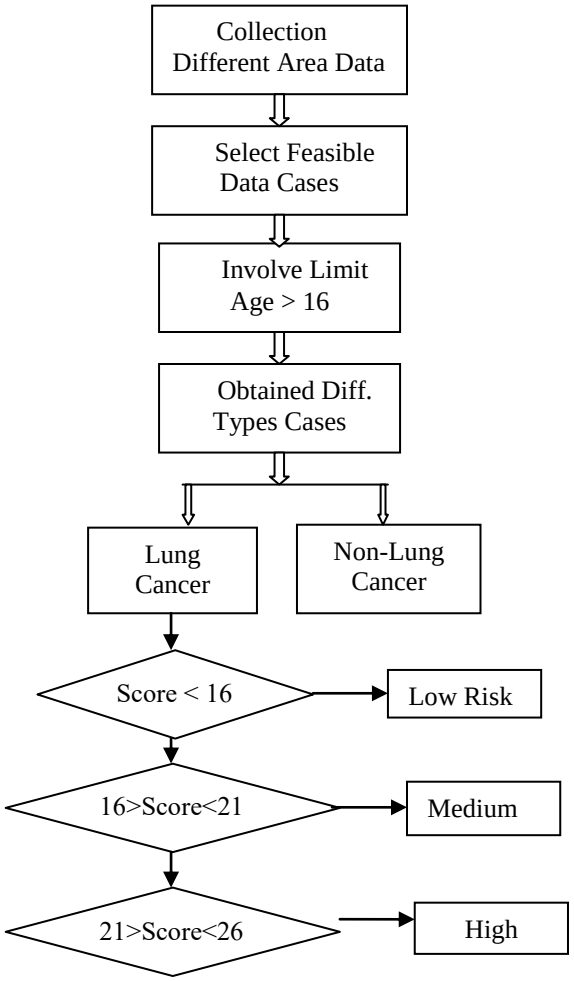

Fig: 1 Risk Diagonic methodology

In this research case study the detection approach uses historical patient huge data and transforms the data into the required format for the classifier by the data preprocessing and feature extraction. Data of lung cancer patients was collected from all different hospitals with various age. These data contains patient-specific information and disease status check below < 15 age and based collection information should matches/ symptom then the particular patent suffer with lung cancer otherwise the patient not cancer diseases. If patient risk factor considers age and dieses symptoms based on again sub categories of disease risk factor (Low Risk, Medium Risk and High Risk).

1. **Low Risk Factor:** A risk analysis consider based on age and disease symptoms, if condition is Score < 16 then primary level consider Low level risk factor. It is helps to doctor take suitable predication on diagonic diseases.
2. **Medium Risk Factor:** A risk analysis, if condition between criteria is Score > 16 <=21then medium level risk factor. It helps to doctor take suitable prediction on diagonic disease.
3. **High Risk Factor:** A risk analysis, if condition between criteria is Score > 21 <=26 then high level risk factor. It helps to doctor take suitable prediction on diagonic disease.

Proper resources for cancer prevention and improper incorporation of correct and sufficient risk factors impacts the early detection and control of this condition. A much better diagnosis is important for timely prevention and control and use of knowledge extraction techniques only can help in tackling with the massive generation and collection of data

ISSN (Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology
An ISO 3297: 2007 Certified Organization Volume 8, Special Issue 1, March 2019
7th National Conference on Frontiers in Communication and Signal Processing Systems (NCFCSPS '19)
19th-20th March 2019
Organized by
Department of ECE, Adhiyamaan College of Engineering, Hosur, Tamilnadu, India

at all stages. The most important input to a well-performing prediction model for identifying cancer and estimating survivability is the selection of correct features.

III. METHODOLOGY

In this research case study the detection approach uses historical patient huge data and transforms the data into the required format for the classifier by the data preprocessing and feature extraction. We integrated multidisciplinary risk factors, including behavioral risk factors, disease history factors, environmental factors, and demographic factors and then preprocessed these integrated data. We have explored and plan to trained data-mining different techniques models in a stratified elderly population. We then extracted risk factors of lung cancer in the elderly and conducted quantitative analyses of the risk factors.

RISK FACTORS

- smoking (cigarette, pipe, and cigar)
- Radon (A naturally occurring gas)
- Asbestos (natural mineral used in construction)
- Others (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 person at greater risk

Fig 2: Risk Factor Phases

1. Smoking Cigarettes

Tobacco use is one of the most important preventable causes of premature death in the world. Smoking harms almost parts of body and increases risk develop of many diseases. Smoking causes most lung cancers and can cause of cancer almost of anywhere on the body. Major causes by smoking as follows:

- Lung cancer
- Breathing problems and chronic respiratory conditions
- Heart diseases, stroke and blood circulation problems

ISSN (Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology
An ISO 3297: 2007 Certified Organization Volume 8, Special Issue 1, March 2019
7th National Conference on Frontiers in Communication and Signal Processing Systems (NCFCSPS '19)
19th-20th March 2019
Organized by
Department of ECE, Adhiyamaan College of Engineering, Hosur, Tamilnadu, India

- Increase risk of miscarriage and irregular periods for women
- Osteoporosis and menopause

Prevention:

- Ways to help people quit smoking and stay tobacco free through counseling, nicotine replacement and other medicine
- Initial college student conducting workshop and awareness of smoking. Better explain details case study of and non-smoking further age condition environment.

2. Radon

Radon exposure is the second leading cause of lung cancer. Radom is a radioactive gas released from the normal decay of the elements uranium, thorium and radium in rocks and soil. It is an invisible, odorless, tasteless gas that seeps up through the ground and diffuses into the air. Radon decays quickly, giving off tiny radioactive particles. When inhaled, these radioactive particles can damage the cells that line the lung, long term exposure to radon can lead to lung cancer. So, suggestion of increased risk of leukemia associated with radon exposure in adults and children. It is major cause by radon as follows

- Radon exposure soil affecting drinking water and weather
- Radon exposure in miners induces gene mutations and chromosomal aberrations.
- Genetic and cytogenetic effects of indoor

Radon symptoms includes shortness of breath, cough, pain or tightness in the chest, hoarseness or trouble swallowing. Radon treat cancer produce it themselves by pumping radon from a radium source and sealing it in small tubes called seeds. The seeds were injected at or near the site of the tumor.

Prevention

- Stop smoking and discourage smoking in your home
- Increase air flow in your house by opening windows and using fans and vents to circulate air
- Seal cracks in floors and walls with plaster, caulk or other materials designer

3. Asbestos

Asbestos is a generic term, which describes a group of diverse, naturally occurring, fibrous minerals. Asbestos exposure is the two main forms of lung cancer are small cell and non-small cell. The symptoms include shortness of breath, chest pain and coughing up blood. Asbestos related lung cancer typically takes between 15 and 35 year to develop from initial exposure to onset of symptoms. Radon symptoms includes shortness of breath, cough, pain or tightness in the chest, hoarseness or trouble swallowing. Radon treat cancer produce, it themselves by pumping radon from a radium source and sealing it in small tubes called seeds. The seeds were injected at or near the site of the tumor. Because of this long latency period and most risk professional involve mining, construction, heavy industry, shipbuilding and firefighting. Major cases symptoms as follows

- Chest pain, weigh loss
- Coughing up Phelgum or sputum
- Difficulty breathing or shortness of breath
- Persistent cough and Fatigue

4. Environmental Tobacco Exposure

Environmental tobacco smoke(ETS) occurs in homes, at workplaces and public places. Environmental tobacco exposure potential health effects regarding the incidence, prevalence and exacerbation. Tobacco smoke contains many chemical effect health like respiratory irritants, systemic toxicant and reproductive toxicants. It major issues as follows

- Reduced lung functionality
- Chronic middle-ear effusion in young children
- Asthma patients increased frequency attacks in children age
- Lower respiratory tract infections

ISSN (Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology
An ISO 3297: 2007 Certified Organization Volume 8, Special Issue 1, March 2019
7th National Conference on Frontiers in Communication and Signal Processing Systems (NCFCSPS '19)
19th-20th March 2019
Organized by
Department of ECE, Adhiyamaan College of Engineering, Hosur, Tamilnadu, India

5. Genetics

A positive family history of lung cancer has been found to be a genetic risk factor influence by pulmonary function. It is major found adjustment age by body habits, cigarette smoking. Increased relative risks were found even after careful adjustment for smoking. Cigarette smoke contain thousand chemicals are modify DNA and lead to the formation of mutations. DNA repairs activities may be risk factors for lung cancer.

Genetic polymorphisms : it is identify multiple genetic polymorphisms underlying lung cancer risk by utilizing up to a million tagging single-nucleotide polymorphisms (SNP) to identify common genetic variations. Polymorphisms in DNA repair genes may be associated with differences in the DRC of DNA damage and may influence an individual's risk of lung cancer, because the variant genotype in those polymorphisms might destroy or alter repair functions.

6. Other Lung Diseases

Oestrogen and progesterone receptors are expressed in the normal lung and in lung cancer cell lines, oestradiol has a proliferative effect on the latter type of cells. A small increased risk of lung cancer has been reported in early studies, while a decreased risk was detected in the more recent studies. No effect was observed in the only randomized trial. While the different results might be explained by changes in the formulations used for replacement therapy, the lack of an effect in the only study with an experimental design argues towards residual confounding by smoking and hence against an effect of this type of exposure on lung cancer.

There is some evidence that a reduced body mass index is associated with an increased risk of lung cancer. However, this inverse association can be explained, at least in part, by negative confounding by smoking and tobacco-related lung disease and no clear association has been demonstrated among never-smokers.

7. Priori Radiation

Exposure to ionising radiation increases the risk of lung cancer. This increased risk has been reported in atomic bomb survivors, as well as patients treated with radiotherapy (RR 1.5–2 for cumulative exposure in excess of 100 cGy). Underground miners exposed to radioactive radon and its decay products, which emit α-particles, have been consistently found to be at increased risk of lung cancer.

There was also evidence that smoking synergistically modifies the carcinogenic effect of radon. Today the main concern about lung cancer risk from radon and its decay products comes from residential rather than occupational exposure.

IV. CONCLUSIONS

Lung cancer prevention, control of tobacco smoking is the most important preventive measure. While the effects of tobacco control in the past few decades on the incidence and mortality of the disease can be appreciated, much remains to be done, in particular among women and in the area of lung cancer screening in smokers using low-dose computed tomography scans. Priorities for the prevention of lung cancer include control of occupational exposures, as well as indoor and outdoor air pollution, and understanding the carcinogenic and preventive effects of dietary and other lifestyle factors to help physicians make decision on cancer

REFERENCES

[1] Kaur S, Bawa RK. Future trends of data mining in predicting the various diseases in medical healthcare system. Int J Energy Inf Commun 2005;6:17-34.
[2] Park SK, Cho LY, Yang JJ, Park B, Chang SH, Lee K, Scientific Committee, Korean Academy of Tuberculosis and Respiratory Diseases. Lung cancer risk and cigarette smoking, lung tuberculosis according to histologic type and Gender in a population based case-control study. Lung Cancer 2010 Apr;68(1):20-26.
[3] Zahnd WE, Eberth JM. Lung cancer screening utilization: a behavioral risk factor surveillance system analysis. Am J Prev Med 2019 Aug;57(2):250-255.
[4] Chen, Hui, Yan Xu, Yujing Ma, and Binrong Ma. "Neural network ensemble-based computer-aided diagnosis for differentiation of lung nodules on CT images: clinical evaluation." Academic radiology 17, no. 5 (2010): 595-602.

ISSN (Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology
An ISO 3297: 2007 Certified Organization Volume 8, Special Issue 1, March 2019
7th National Conference on Frontiers in Communication and Signal Processing Systems (NCFCSPS '19)
19th-20th March 2019
Organized by
Department of ECE, Adhiyamaan College of Engineering, Hosur, Tamilnadu, India

[5] Jiawei Han and Micheline Kamber, "Data Mining Concepts and Techniques", Second Edition, University of Illinois at Urbana-Champaign, 2006.
[6] Songjing Chen, PhD; Sizhu Wu, PhD, "Identifying Lung Cancer Risk Factors in the Elderly Using Deep Neural Networks: Quantitative Analysis of Web- Based Survey Data" Journal of Medical Internet Research, Med Internet Res 2020 | vol. 22 | iss. 3 | e17695 |
[7] Ada, Early Detection And Prediction Of Lung Cancer Survival Using Neural Network Classifier, International Journal Of Application Of Innovation In Engineering Of Management (Ijaiem), Volume 2, Issue 6, June 2013
[8] Ramachandran P, Girija N, Bhuvaneswari T. Early detection and prevention of cancer using data mining techniques. Int J Comput Appl 2014;97:48-53.
[9] 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.
[10] 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.
[11] Park SK, Cho LY, Yang JJ, Park B, Chang SH, Lee K, Scientific Committee, Korean Academy of Tuberculosis and Respiratory Diseases. Lung cancer risk and cigarette smoking, lung tuberculosis according to histologic type gender in a population based case-control study. Lung Cancer 2010 Apr;68(1):20-26. [doi: 10.1016/j.lungcan.2009.05.017]
[12] A. A. Freitas, "A Genetic Programming Framework for Two Data Mining Tasks: Classification and Generalized Rule Induction," 1997
[13] P.Thangaraju , G.Barkavi, "Lung Cancer Early Diagnosis Using Some Data Mining Classification Techniques: A Survey" COMPUSOFT, An international journal of advanced computer technology(IJACT), 3 (6), June-2014 (Volume-III, Issue-VI)
[14] World Health Organization. WHO Statistical Information System (WHOSIS). www.who.int/healthinfo/mortality_ data/en/ Date last updated: November 25, 2015. Date last accessed: December 15, 2015.
[15] Zakaria Suliman Zubi, Rema Asheibani Saad, "Using Some Data Mining Techniques for Early Diagnosis of Lung Cancer, ACM DL, RecentResearch in Artificial Intelligence, KnowledgeEngineering and Data Bases, 2011, 32-37.
[16] Krishnaiah V, Narsimha G, and Chandra NS, "Diagnosis of lung cancer prediction system using data mining classification techniques," International Journal of Computer Science and Information Technologies, vol. 4, pp. 39–45, 2013.
[17] Yongqian Qiang, Youmin Guo, Xue Li, Qiuping Wang, Hao Chen, & Duwu Cuic 2007 .The Diagnostic Rules of Peripheral Lung cancer Preliminary study based on Data Mining Technique. Journal of Nanjing Medical University, 21(3):190-195