# A Study on Mining Lung Cancer Data for Increasing or Decreasing Disease Prediction Value by Using Ant Colony Optimization Techniques

**Dr.T.Christopher,**
**Assistant Professor, Department of Computer Science, Government Arts College,Coimbatore,**
**chris.hodcs@gmail.com**

**J**.**Jamera banu,**
**M.Phil., Scholar, Department of Computer Science, Government Arts College,Udumalpet,**
**jameramphil@gmail.com**

-------------------------------------------------------------ABSTRACT-------------------------------------------------------------
Cancer is the most Important Cause of death for both men and women. Lung Cancer is the Uncontrolled growth of abnormal cells that start off in one or both Lung Usually in the cell that Line the air Passages. The Two main types are Small Cell Lung Cancer and Non-Small Cell Lung Cancer. 10-15% of Lung Cancer Occur in Non-Smokers. 50% is Occur in Smokers. The Risk of Lung Cancer Increase with the length of time and number of Cigarettes they have Smoked. Lung Cancer has Spread, Person may feel Symptoms Such as Age, Sex, Wheezing, Shortness of Breath, Chest Pain it can predict the like hood of patient getting Lung Cancer Disease. Data mining Techniques used to Detect Early Stage of Lung Cancer Disease Using Data mining Algorithm Such as Data mining, Lung Cancer Prediction, Classification, Decision Table, Naïve Based, Ant Colony Optimization. Study of this Paper the Early Detection of Lung Cancer can be curing the Disease Completely it helpful in doctor saving the life of the patients. Data mining Techniques of Ant Colony Optimization is helps Lung Cancer Data for Increasing or Decreasing Disease Prediction Value.

**Keywords**
        ACO (Ant Colony Optimization), Classification, Data mining, Decision Table, Lung Cancer Prediction.
-------------------------------------------------------------------------------------------------------------------------------- ----------

## 1.INTRODUCTION

Lung cancer is a cancer that starts in the lungs. Lung cancer is the most common cause of cancer death worldwide. The occurrence of lung cancer has increased rapidly and become the most common cancer in men in most countries. Smoking is by far the most important preventable cause of cancer in the world. If the original lung cancer has spread, a person may feel symptoms in other places in the body. Common places for lung cancer to spread include other parts of the lungs in Tumor, lymph nodes, bones, brain, liver. The incidence of lung cancer is strongly correlated with cigarette smoking, with about 90% of lung cancers arising as a result of tobacco use. The risk of lung cancer increases with the number of cigarettes smoked over time. The risk of lung cancer increases with the number of cigarettes smoked over time. Most people know that smoking causes cancer, but may not realize how many nonsmokers get lung cancer, too. The purposes of this work is finding the risk factor of lung cancer and classify the smokers and non-smokers who are all caused by lung cancer by using the data mining Technique.

## 2. RELATED WORKS
Thangaraju P, et al.,[1] Mining Lung Cancer Data for Smokers and Non-Smokers by Using Data Mining Techniques Lung cancer is a cancer that starts in the lungs. Smoking is the biggest risk factor of lung cancer. The more years and larger number of cigarettes smoked the greater the risk of developing lung cancer. The average age of someone diagnosed with lung cancer is 65 to 70 years old, but people who are younger can develop lung cancer. Young adults who have never smoked also can develop lung cancer. This Paper is to finding the risk factor of lung cancer. It is hoped on prevention of lung cancer for people.

Krishnaiah V, et al.,[2] proposed to a model for nearly detection and correct diagnosis of the disease which will help the doctor in saving the life of the patient. Using generic lung cancer symptoms such as age, sex, Wheezing, Shortness of breath, Pain in shoulder, chest, arm, it can predict the likelihood of patients getting a lung cancer disease.

ParagDeoskar, et al.,[3] proposed to assorted data mining and ant colony optimization techniques for appropriate rule generation and classification, which pilot to exact cancer classification. In addition to, it provides basic framework for further improvement in medical diagnosis. This paper also survey the aspects of ant colony optimization (ACO) technique. Ant colony optimization helps in increasing or decreasing the disease prediction value.

Sowmiya T, et al., [4] Lung cancer is one of the most dangerous cancer types in the world. These diseases can spread worldwide by uncontrolled cell growth in the tissues of the lung. Early detection of the cancer can save the life and survivability of the patients who affected by this diseases. In this paper we survey several aspects of data mining procedures which are used for lung cancer prediction for the patients. Data mining concepts is useful in lung cancer classification. We also reviewed the aspects of ant colony optimization (ACO) technique in data mining. Ant colony optimization helps in increasing or decreasing the disease prediction value of the diseases. This case study assorted data mining and ant colony optimization techniques for appropriate rule generation and classifications on diseases, which pilot to exact Lung cancer classifications. In additionally to, it provides basic framework for further improvement in medical diagnosis on lung cancer.

Prashant Naresh et al.,[5] proposed to proposed to significant pattern prediction tools for a lung cancer prediction system were developed. The lung cancer risk prediction system should prove helpful in detection of a person's predisposition for lung cancer. The early prediction of lung cancer should play a pivotal role in the diagnosis process and for an effective preventive strategy.

## 3. DATA MINING TECHNIQUE

Data mining is the process of automatically collecting large volumes of data with the objective of finding hidden patterns and analyzing the relationships between numerous types of data to develop predictive models. The classification techniques 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. Such analysis can help provide us with a better understanding of the data at large.

## 4. DATA SET

Dataset used in this model should be more precise and accurate in order to improve the predictive accuracy of data mining algorithms. Which is collected may have missing (or) irrelevant attributes. These are to be handled efficiently to obtain the optimal outcome from the data mining process. Age, Gender, Height, Weight, Smoking habit, Secondhand smoke, Radon gas, Asbestos, Air pollution, Radiation therapy to lungs, HIV or AIDS, Organ Transplant.

In the method mainly decision tree is used for predicting the Lung Cancer Disease from the given data set instances and the proposed model contains three different types of decision tree algorithms such as Naïve Bayes, Decision Table and j48 are applied on type Lung Cancer Disease dataset in the WEKA tool and the performance is calculated. Here the framework can be given as below and the performance can be obtained based on the time taken to build the tree and correctly classified instances.

The time taken by the algorithms(Naive Bayes, Decision Table, and J48) to build the decision tree in the weka tool. In the above table the time is mentioned in Mille seconds. The Naive Bayes takes 0.01 ms for build the decision tree and the Decision Table takes 0.05 ms, at the same time J48 takes 0.03 ms for build the decision tree in the weka tool. By considering the above table we can easily say the Naive Bayes algorithm is the best performance algorithm based on the time. The dataset consists of 303 instances and they are applied as a test case in the classification algorithms. The performance of the algorithms can be known from the instances that are correctly classified. Each algorithm classifies different number of instances. The instances which are correctly classified using the WEKA tool can be given as below.

Accuracy measures the ability of the classifier to correctly classify unlabelled data.

$$\text{Accuracy} = \frac{\text{Number of objects correctly Classified}}{\text{Total No. of objects in the test set.}}$$

Accuracy measure is correctly classified data. In this work, Classification method is used to analyze the Somker's and non smoker's risk factors based on each cells of human and the stages of lung cancer with the help of Weka tool. Which will helps to find out the complications in lung cancer and treat earlier.

## 5. CLASSIFICATION OF LUNG CANCER

Classification method is used to analyze the Somker's and non smoker's risk factors based on each cells of human and the stages of lung cancer with the help of Weka tool.
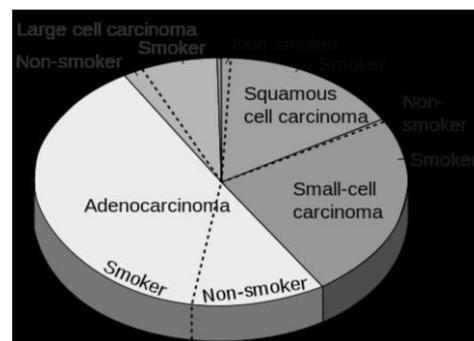


Classification of lung cancer based on cell carcinoma.

### 5.1 Adenocarcinoma

Adenocarcinoma is a common histological form of lung cancer. Nearly 40% of lung cancers are adenocarcinoma, which usually originates in peripheral lung tissue. Most cases of adenocarcinoma are associated with smoking; however, among people who have smoked fewer than 100 cigarettes in their lifetimes("never-smokers"), adenocarcinoma is the most common form of lung cancer.

## 5.2 Squamous cell carcinoma

These cancers start in early versions of squamous cells, which are flat cells that line the inside of the airways in the lungs. They are often linked to a history of smoking and tend to be found in the middle of the lungs, near a bronchus.

## 5.3 Large cell carcinoma

This type of cancer accounts for about 10% to 15% of lung cancers. It tends to grow and spread quickly, which can make it harder to treat. A subtype of large cell carcinoma, known as large cell neuroendocrine carcinoma, is a fast growing cancer that is very similar to small cell lung cancer.

## 5.4 Small cell carcinoma

Small cell carcinoma often starts in the bronchi near the center of the chest, and it tends to spread widely through the body fairly early in the course of the disease.

## 6. PRE-DIAGNOSIS TECHNIQUES

Pre-diagnosis helps to identify or narrow down the possibility of screening for lung cancer disease. Symptoms and risk factors (smoking, alcohol consumption, obesity, and insulin resistance) had a statistically significant effect in pre-diagnosis stage. The lung cancer diagnostic and prognostic problems are mainly in the scope of the widely discussed classification problems. These problems have attracted many researchers in computational intelligence, data mining, and statistics fields. 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 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. The objective of this study is to summarize various review and technical articles on diagnosis of lung cancer. It gives an overview of the current research being carried out on various lung cancer datasets using the data mining techniques to enhance the lung cancer diagnosis.

Data mining Technique of prediction technique is based on systematic study of the statistical factors, symptoms and risk factors associated with Lung cancer. Non-clinical symptoms and risk factors are some of the generic indicators of the cancer diseases. Initially the parameters for the pre-diagnosis are collected by interacting with the pathological, clinical and medical oncologists (Domain experts).

## 7. LUNG CANCER SYMPTOMS

The following are the generic lung cancer symptoms
- Coughing up blood (heamoptysis) or bloody mucus.
- Chest, shoulder, or back pain that doesn't go away and often is made worse by deep Hoarseness
- Weight loss and loss of appetite
- Increase in volume of sputum
- Wheezing
- . Shortness of breath
- Repeated respiratory infections, such as bronchitis or pneumonia
- Repeated problems with pneumonia or bronchitis
- Fatigue and weakness
- New onset of wheezing
- Swelling of the neck and face
- Clubbing of the fingers and toes. The nails appear to bulge out more than normal.
- Paraneoplastic syndromes which are caused by biologically active substances that are secreted by the tumor.
- Fever
- Hoarseness of voice
- Puffiness of face
- Loss of appetite
- Nausea and vomiting

## 8. LUNG CANCER RISK FACTORS

Lung Cancer is affected in many Risk Factors. The Risk Factors is following below. Such as,
- Smoking: Beedi, Cigarette and Hukka
- Second-hand smoke
- Radon exposure
- Air pollution
- Insufficient consumption of fruits & vegetables
- Suffering with other types of malignancy.

## 8. EXISTING METHOD

Lung cancer is the number one cause of cancer deaths in both men and women in the U.S. and worldwide. Cigarette smoking is the principal risk factor for development of lung cancer. The stage of lung cancer refers to the extent to which the cancer has spread in the body. Overall, 10-15% of lung cancers occur in non-smokers. (Another 50% occur in former smokers). Two-thirds of the non-smokers who get lung cancer are women and 20% of lung cancers in women occur in individuals who have never smoked. Lung Cancer is the most critical reason for death.

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 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. The future enhancement can be applicable for image based analysis for MRI and CT scan images.

## 9. CONCLUSION

In this paper Data mining plays a major role in extracting the hidden information in the medical data base. The data preprocessing is used in order to improve the quality of the data. This paper test for dataset and successfully performed with several data mining classification techniques. 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. In some cases even in the advanced level Lung cancer patients does not show the symptoms associated with the Lung cancer. 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 that the Prevalence and mortality rate could be brought down. Study of this paper develop an The accuracy classification Techniques used in several dataset is help in Increasing and Decreasing Disease improve Prediction value and improve lack of awareness help in the Lung Cancer patients, Research and ultimately improve the quality of health care of Lung Cancer patients.

## REFERENCES

[1] Thangaraju P, Karthikeyan T, Barkavi G, Mining Lung Cancer Data for Smokers and Non-Smokers by Using Data Mining Techniques, *International Journal of Advanced Research in Computer and Communication Engineering Vol. 3, Issue 7, July 2014.*

[2] Krishnaiah V, Narsimha G, Subhash Chandra N , Diagnosis of Lung Cancer Prediction System Using Data Mining Classification Techniques, *International Journal of Computer Science and Information Technologies, Vol. 4 (1) , 2013, 39 – 45.*

[3] Parag Deoskar, Dr. Divakar Singh, Dr. Anju Singh, Mining Lung Cancer Data And Other Diseases Data using Data Mining Techniques, *A Survey, Volume 4, Issue 2, March – April (2013).*

[4] Sowmiya T, Gopi M, Thomas Robinson, Optimization of Lung Cancer using Modern Data Mining Techniques, *International Journal of Engineering Research Volume No.3, Issue No.5.*

[5] Prashant Naresh,, Early Detection of Lung Cancer Using Neural Network Techniques, *Prashant Naresh Int. Journal of Engineering Research and Applications www.ijera.com ISSN : 2248-9622, Vol. 4, Issue 8( Version 4), August 2014, pp.78-83.*