Research Paper

(Term: Jan-May 2022)

<u>Lung Cancer Predictions</u>

Submitted By
Name: Abhishek Kumar Singh
Reg. No. -11907788

Under the Guidance of
Dr. Sagar Pande
<u>School of Computer Science and Engineering</u>
LOVELY PROFESSIONAL UNIVERSITY
PHAGWARA, PUNJAB

Abstract

Cancer is the most important cause of death for both men and women. The early detection of cancer can be helpful in curing the disease completely. So the requirement of techniques to detect the occurrence of cancer nodule in early stage is increasing. Earlier diagnosis of Lung Cancer saves enormous lives, failing which may lead to other severe problems causing sudden fatal end. Machine learning is a powerful technique to help the people in their health, Scientific and Engineering. Those techniques are extracting the hidden information from the large databases which helps to find the relationships and patterns from the data. In this prediction we use different algorithms to predict lung Cancer using different Model .

Introduction

Lung cancer is the one of the leading cause of cancer deaths in both men and women. Manifestation of Lung cancer in the body of the patient reveals through early symptoms in most of the cases . Treatment and diagnosis depend on the histological type of cancer, the stage (degree of spread), and the patient's performance status. Possible treatments include chemotherapy, and radio therapy , surgery Survival depends on stage, overall health, and other factors, but overall only fourteen percent of people diagnosed with lung cancer survive five years after the diagnosis. In simple words cancer is abnormal growth of cells, and ultimately to the stoppage of the essential cellular functions of the organism. These cells are generally called tumor cells and they often clump together into lumps to form 'Tumors'.

Early symptoms may include-

- Bad painful cough
- Cough with phlegm or blood
- Heavy chest pain while deep breathing, laughing, coughing or any other muscular activities near chest
- shortness of breath
- weakness and fatigue
- loss of appetite and weight loss

Causes - About ninety percent of lung cancer cases involve smoking. Smoking causes destruction of lung tissues, lung can repair them but heavy smoking makes lung stopping it's natural behaviour . A radioactive gas Radon, is the second leading cause, according to the American Lung Association. Breathing in other hazardous substances can also cause lung cancer if it happens over a long period of time. A type of lung cancer called mesothelioma is almost always caused by exposure to asbestos.

Objective and Dataset

The main objective of the project is to construct a program used for the predictions of Lung cancer using python machine learning. I used Jupyter Notebook as my working platform or IDE during this Project . And the Dataset was downloaded from Kaggle . Different libraries and modules has been used in the process.

Methodology

In this project I try to predict lung cancer using different algorithms.

Predictor variable use in classifying lung cancer:- Age, Smokes , AreaQ , Alkhol

Step-

- First I loaded the data set after that import the data
- Creating data comparison between predict value using graphs
- Split the dataset and train the data
- Predict the dataset using different algorithms
- Calculate the accuracy for all algorithms
- Find the best model for Dataset

Model

- **Logistic Regression-** Logistic regression is a technique that can be applied to binary classification problems**. Its gives the accuracy of 82% .**

- **Support Vector Machine-** it is a supervised machine learning algorithm that can be employed for both classification and regression purposes**. Its gives the accuracy of 81.5% .**

- **K-Nearest Neighbor classification-** It belongs to the supervised learning domain and finds intense application in pattern recognition, data mining and intrusion detection. **Its gives the accuracy of 86.5% .**

- **Decision tree classification-** The goal of decision tree is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features. **Its gives the accuracy of 71% .**

Conclusion

Lung cancer is one of the most dangerous diseases and the most common cause of death, the severity of the disease lies in the difficulty of diagnosing it in the early stages. This paper tries to endeavor to investigate of four classifiers to find the best classifier could classify lung cancer in early stage. A prototype lung cancer disease prediction system is developed using different classifier algorithms. For example Logistic Regression, Support Vector Machine , K- Neighbour Classification algorithms and Decision tree algorithms. Lung cancer prediction system can be further enhanced and expanded. It can also incorporate other data mining techniques, e.g., Time Series, Clustering and Association Rules.