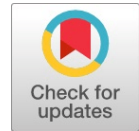


# Lung Cancer Prognosis using Machine Learning



Lavi Vashishth, Kanishka Jain, Aditya Maheshwari, Prasun Chakrabarti

**Abstract:** The paper calls attention to investigation of Lung Cancer in the light of regulated AI classifiers. The parameters incorporate age, smokes, passive smoking, alcohol and gender.

Curve fitting has likewise been applied in request to break down the Lung Cancer. These days, around 33% of grown-ups are known to be smokers, and smoking rates are expanding among the female population. Tobacco use has been accounted for to be the fundamental driver of 90% of male and 79% of female lung malignancies. 90% of passing from lung malignancy are evaluated to be because of smoking.

**Keywords :** malignancy, age, smoke, tobacco, alcohol.

## I. INTRODUCTION

Lung malignancy is the main source of Cancer passing in the two people in the U.S. What's more, Cigarette smoking is the chief hazard factor for advancement of lung malignant growth. Lung disease is the main source of long periods of life lost due to malignant growth and is related with the most noteworthy financial weight comparative with other tumour types. Research stays at the foundation of accomplishing improved results of lung malignancy.

Lung disease, otherwise called **lung carcinoma**, is a dangerous lung tumour portrayed by uncontrolled cell development in tissues of the lung.

Introduction to radon, a normally existing radioactive gas, is the subsequent driving reason, as indicated by the American Lung Association. Radon is a gas that happens normally with the breakdown of uranium in rocks and soil. Radon is hard to recognize and could be uncovered without knowing it. Radon is the second-driving reason for lung disease in the United States.

### Causes

Anybody can get lung malignancy, yet 90 percent of lung disease cases are the after effect of smoking.

Smoking tobacco is the main hazard factor for lung disease, representing about 90 percent everything being equal, as per the CDC. Tobacco and tobacco smoke contain in excess of 7,000 synthetic substances, a considerable lot of which are cancer-causing like nitrogen oxide and carbon monoxide.

**Manuscript published on 30 March 2020.**

\* Correspondence Author (s)

**Lavi Vashishth**, Department of Computer Science and Engineering, Techno India NJR Institute of Technology Udaipur, Rajasthan, India.

**Kanishka Jain**, Department of Computer Science and Engineering, Techno India NJR Institute of Technology Udaipur, Rajasthan, India.

**Aditya Maheshwari**, Capacity of Project Lead- New Initiative & Research, Techno India NJR Institute of Technology Udaipur, India.

**Prof Dr Prasun Chakrabarti**, Executive Dean (Research and International Linkage) and Institute Distinguished Senior Chair Professor, Techno India NJR Institute of Technology, Rajasthan, India.

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](https://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC-BY-NC-ND license <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

### Manifestations include:

- craving misfortune
- visit chest diseases, for example, bronchitis or pneumonia
- brevity of breath
- unexplained migraines
- weight reduction

### Organizing:

The different phases of lung malignant growth are as per the following: ,

**Stage 0:** The specialist finds unusual cells just in the top layers of cells covering the aviation routes.

**Stage I:** A tumour has created in the lung, however is under 5 centimetres (cm) and has not spread to different pieces of the body.

**Stage II:** The tumour is littler than 5 cm and might have spread to the lymph hubs in the territory of the lung, or littler than 7cm.

**Stage III:** Cancer has spread to the lymph hubs and arrived at different pieces of the lung and encompassing zone.

**Stage IV:** Cancer has spread to removed body parts, for example, the bones or cerebrum.

### Treatment:

Potential medications of lung malignancy include: Surgery, Chemotherapy. Chemotherapy treatment utilizes medications to contract or annihilate disease cells.

## II. METHODOLOGY

The dataset for the venture is taken from the UCI Machine Learning Repository.

<https://www.kaggle.com/yusufdede/lung-cancer-dataset>

Dataset involves 59 instances and 6 attributes. The yield depends on results on the way that how much populace including males and females is brittlely influenced by Lung Cancer.

For our examination we have utilized Weka 3.8.4 apparatus to characterize information on Lung Cancer. Based on region we have mapped the qualities for smokes, passive smoking, ages and alcohol. Likewise, Microsoft Excel has been utilized for graphical portrayal of the equivalent.

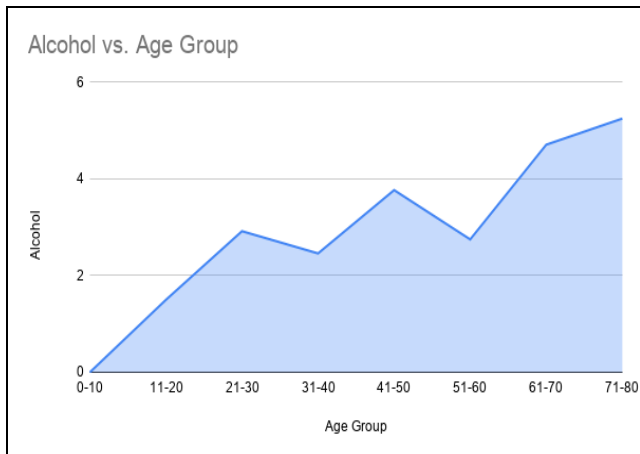
## III. RESULTS AND ANALYSIS

# Lung Cancer Prognosis using Machine Learning

**Table 1: Weka based analysis of lung cancer**

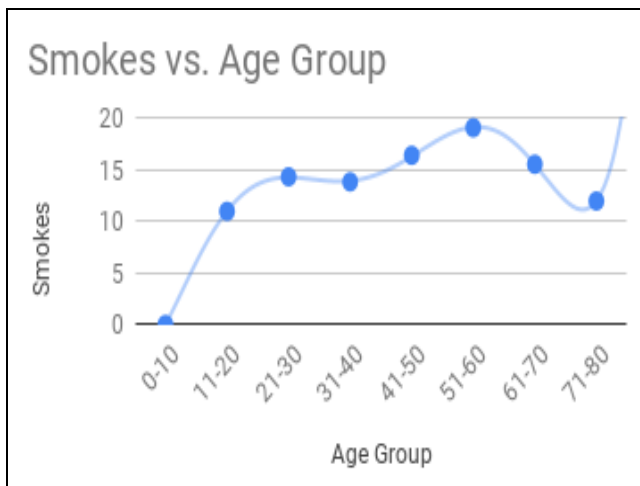
Classifiers	Correlation Coefficient	Mean Absolute Error	Root Mean Squared Error	Relative Absolute Error	Root Relative Squared Error
lazy.IBK	0.934	0.0339	0.1841	0.066	0.36
functions.LinearRegression	0.9174	0.1499	0.199	0.293	0.389
lazy.Kstar	0.9433	0.0494	0.167	0.096	0.326
trees.RandomForest	0.9247	0.1063	0.1937	0.208	0.378
functions.SMOreg	0.9241	0.1437	0.1919	0.281	0.375
functions.MultilayerPerceptron	0.9142	0.0816	0.2086	0.159	0.407
trees.REPTree	0.7702	0.1541	0.3247	0.301	0.634
trees.RandomTree	0.7629	0.1186	0.3444	0.232	0.673
misc.InputMappedClassifier	-0.4523	0.5102	0.5114	0.01	0.01

We have used weka tool in order to find out accuracy estimate of lung cancer prognosis in the light of supervised machine learning classifiers. From the experimental results it is evident that Lazy KStar generates better correlation coefficient (94.33%).



**Figure 1: Shows the graph plot between the age group of people vs. the alcohol consumption by each of the age groups**

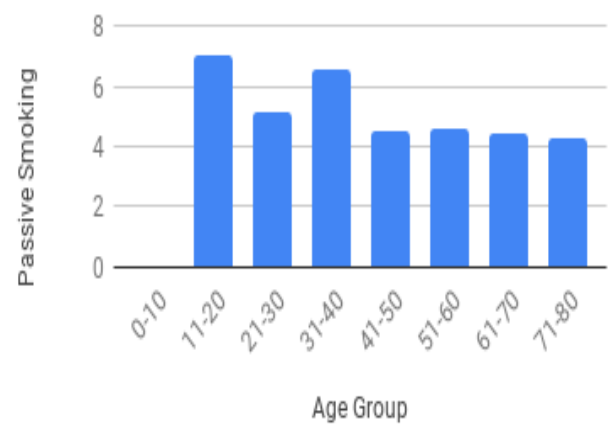
From Figure 1 it is clear that the age group (71-80) has been affected mostly by lung cancer due to various factors



**Figure 2: Age wise distribution of persons affected by lung cancer due to smoking.**

Here, it is clear that age group 51-60 years is highly affected by smoking.

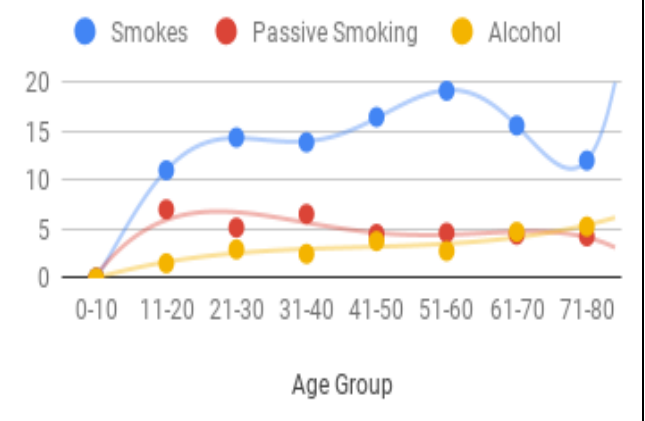
## Passive Smoking vs. Age Group



**Figure 3: Age wise distribution of persons affected by lung cancer due to passive smoking.**

Here, it is clear that age group 11-20 years is highly effected by passive smoking.

## Smokes, Passive Smoking and Alcohol



**Figure 4: Graph Plot between various Age Groups vs. the factors which causes lung cancer i.e. Smoke, Passive Smoking and Alcohol.**

This above figure shows that all groups are affected by smoking more than passive smoking or alcohol.

#### IV. CONCLUSIONS

The paper points out a technique based on Weka tool and available dataset in context to Lungs Cancer. The significant factors contributing the cause of Lungs Cancer include smoking, passive smoking and alcohol.

On the basis of Supervised Machine Learning classifiers, it is observed that

1. Lazy Kstar generates better accuracy estimates (94.33%).
2. From the graphical representation it is evident that person under the age group 51-60 are affected by addiction towards smoking, 11-20 by tobacco and 71-80 by alcohol.

This study will facilitate the doctors towards prognosis of Lungs Cancer and the weightage of each factor leading to the disease can be noted using supervised learning rule.

#### REFERENCES:

1. *Wikipedia*. Retrived from en.wikipedia.org/wiki/Lung\_cancer
2. Pietrangelo A. (2019). "Everything You Need to Know About Lung Cancer". *Healthline*. May 14, 2019. Retrieved from www.healthline.com/health/lung-cancer#causes
3. Nall R., MSN, CRNA (2018). "What to know about lung cancer". *Medical News Today*. 16 November, 2018. Retrieved from www.medicalnewstoday.com/articles/323701#what-is-lung-cancer.
4. Davis C.P. "Cancer". *MedicineNet*. Retrieved from www.medicinenet.com/cancer/article.htm
5. Stöppler M.C. "Smoking and How to Quit Smoking". *MedicineNet*. Retrieved from www.medicinenet.com/smoking\_and\_quitting\_smoking/article.htm

#### AUTHORS PROFILE



**Lavi Vashishth** is a student of first year, Computer Science and Engineering at Techno India NJR Institute of Technology Udaipur, Rajasthan. Her research interest is mostly based on statistical modeling.



**Kanishka Jain** is a student of first year, Computer Science and Engineering at Techno India NJR Institute of Technology Udaipur, Rajasthan. Her research interest is data analytics and tools.



**Aditya Maheshwari** is working in the capacity of Project Lead- New Initiative & Research at Techno India NJR Institute of Technology Udaipur. He is an Open Source Advocate, leading technical speaker and Expert of User Research (UI/UX).



**Prof Dr Prasun Chakrabarti** has received his PhD (Engg.) from Jadavpur University in 2009. He is working as Executive Dean (Research and International Linkage) and Institute Distinguished Senior Chair Professor, Techno India NJR Institute of Technology.