# Lung Carcinoma - A Prophesies on Survivability and Recurrence Rate

S.Divya Meena
*Student, Department of Conputer Science & Engineering*
*Kingston Engineering College*
*Vellore, India*

M.Revathi
*Asst. Professor, Department of Conputer Science & Engineering*
*Kingston Engineering College*
*Vellore, India*

**Abstract- Lung cancer currently ranks as the leading cause of cancer related deaths in both gender, though it is one of the most preventable forms of cancers. For men, the incidence of lung cancer has declined, but in women it continues to rise. Beyond being the most common form of cancer, it is often difficult to treat. Beyond treating cancer, detecting the presence of cancer is itself difficult, involving several tests. While there are different types of treatment for lung cancer, the choice of treatment should be based on type of cancer. The overall stage of cancer is influenced by the TNM Classification system. The prognosis of lung cancer requires several interesting attributes including Spiculation, Emphysema, etc. Recurrence of lung cancer can be minimized with Post-operative Radiotherapy (PORT). Patient's Survival prediction for lung cancer remains problematical. The aim of the present study is to examine the medical efficiency of an conventional bio-marker as the basis of risk stratification in patients with lung cancer.**

**Keywords – Radiotherapy, TNM Classification System, Lung Carcinoma, Emphysema, Smoking, Spiculation**

## I.INTRODUCTION

Lung cancer survival has not differed ominously for decades, but new multidisciplinary approaches are starting to improve the outcomes. In the contemporary world, there have been more improvements in the care of lung cancer patients than at any other time in recent history. However, it is becoming ever more challenging to stay abreast of all of the advances that evolves from various fortes of lung cancer management. The objective is to predict the incidence of lung cancer, its survivability and recurrence rate. The occurrence of lung cancer can be detected with series of test involving Test for Pulmonary Vascular Resistance and Spirometry initially. Pulmonary Vascular Resistance is the resistance of the pulmonary circulation and Spirometry is a test to diagnose various lung conditions, mainly COPD, Chronic Obstructive Pulmonary Disease. We then predict the probability of Non-small cell lung cancer (NSCLC). It is the most common lung cancer accounting to 80% approximately. The stage of cancer can be estimated with the most common staging system, TNM System developed by the International Union against Cancer. The cancer probability can be predicted with some prognostic factors such as Age, Gender, Node size, Node Type, Nodes count, Presence of nodes in upper lung, Emphysema and Spiculation. The survival benefit of post-operative radiotherapy (PORT) for non-small cell lung cancer (NSCLC) remains controversial. So, we construct a survival prediction model to enable individualized predictions of the net survival benefit of PORT for NSCLC cancer patients based on the tumour characteristics such as Histology, Tumour extent and node extent.

## II.RELATED WORKS

Bounteous amount of work have been done to predict the risk of the cancer. In the research finding, certain traits of our work are compared with previous works. Some of them are as follows:

| **Type** | **Endpoint** | **Algo** | **Data** | **Ref** |
|---|---|---|---|---|
| Lung | Survivability | DT | Mixed | Catto et al |
| Lung | Recurrence | NN | Clinical | Jin et al |
| Lung | Survivability | DT | Mixed | Listgarten et al |

| | | | | |
|---|---|---|---|---|
| Lung | Susceptibility | SVM | Mixed | Delen et al |
| Lung | Survivability | Naïve Bayes | Mixed | Bellaachia et al |
| Lung | Recurrence | NN | Clinical | Grumett et al |
| Lung | Treatments | SVM | Clinical | Hayashida et al |
| Lung | Recurrence | DT | Clinical | Masic et al |
| Lung | Recurrence | ANN | Mixed | Luna et al |
| Lung | Survivability | DT | Clinical | Garcia et al |
| Lung | Survivability & Recuurence | RF DT | Clinical and Gene data | This paper |

## III.METHODS AND TECHNIQUES

*3.1 Feature Selection –*

Feature selection is a pre-processing technique that helps in identifying and removing features that are irrelevant to the classification and in so doing produces a compact data set. This method will increase the accurateness without altering the relevance of the features and so used widely in Healthcare field.

| **Trivial Variables in Breast Cancer** | **Significant features in Breast Cancer** |
|---|---|
| Age | Age & Gender |
| Gender | Node size |
| Cancer grade | Hormone Node Type |
| Tumour size | Nodes count |
| Physical condition | Emphysema |
| Smoking | Spiculation |

*3.2 Prognosis Procedure*

NSCLC is any type of epithelial lung cancer other than small cell lung cancer (SCLC). Familiar types of NSCLC includes squamous cell carcinoma, large cell carcinoma, and adenocarcinoma, but there are quite a few other kinds that transpire less frequently but all types can occur in unusual histologic variants. Although NSCLCs are related with chain-smokers, adenocarcinomas may be found in patients who have never smoked. As a class, NSCLCs are pretty oblivious to chemotherapy and radiation therapy compared with SCLC. Patients with resectable disease may perhaps be alleviated by surgery only or surgery charted with chemotherapy. Local control can be achieved with radiation therapy in a outsized amount of patients with unresectable disease, but cure is seen only in a small number of patients. Patients with locally progressive unresectable syndrome may perhaps attain long-term survival with radiation therapy combined with chemotherapy. Patients with unconventional metastasized disease may attain better survival and palliation of symptoms with chemotherapy, targeted agents, and other supportive measures.

*3.3 Lung Cancer Stages*

Accurately characterizing the anatomic extent of disease in a patient with lung cancer can help with better cure and diagnosis. Non-small cell lung cancer is staged using the TNM system (T for extent of primary tumour, N for regional lymph node involvement, and M for metastasis).

| **TNM** | **DESCRIPTION** |
|---|---|
| T1 | A small tumour that is not locally advanced or invasive |
| T2 | A larger tumour that is minimally advanced or invasive |
| T3 | Any size of tumour that is locally progressive otherwise aggressive up to but not including the major intrathoracic structures |

| | |
|---|---|
| T4 | Any size tumour that is advanced or invasive into the major intrathoracic structures |
| N0 | Metastatic disease to nodes within the Ipsilateral lung |
| N1 | Metastatic disease to nodes beyond the Ipsilateral lung but not contralateral to the primary tumour |
| N2 | Metastatic disease to nodes distant to those included in N2 |
| M0 | Local or regional disease |
| M1 | Disseminated disease |

## IV.PREDICTOR MODEL BUILDING

### *4.1 Pulmonary Vascular Resistance of Lungs*

Pulmonary Vascular Resistance is the resistance of the pulmonary circulation. To find the PVR of lungs, we use three basic factors namely Mean Pulmonary Artery Pressure; Pulmonary Capillary Wedge Pressure and Pulmonary Flow. The normal values of MPAP, PCWP and PF are given as >25mmHg, 4-12 mmHg and 4-8 L/min respectively. With these factors we estimate if PVR is normal or abnormal. The normal rate of PVR is in the range of 37-250.

### *4.2 Spirometry Test for COPD*

Spirometry is a test to diagnose various lung conditions, mainly COPD, Chronic Obstructive Pulmonary Disease. It measures if the value of FVC, FEV1 and PEFR are normal or not. It includes other essential factors such as Age, Gender and Height to check if FEV1/FVC is normal or low. In case the FEV1/FVC value is low, then we should consider COPD.

### *4.3 Non-Small Cell Lung Cancer Probability*

The cancer probability can be predicted with some prognostic factors such as Age, Gender, Node size, Node Type, Nodes count, Presence of nodes in upper lung, Emphysema and Spiculation.

| FACTORS | VALUES | SCORE |
|---|---|---|
| **Age** | <40 | 1 |
| | 40-69 | 2 |
| | >=70 | 3 |
| **Gender** | Male | 1 |
| | Female | 0 |
| **Nodule Type** | Solid | 0 |
| | Non-solid | 1 |
| | Partially Solid | 2 |
| **Nodule size** | 0-9 | 1 |
| | 10-12 | 2 |
| | 20-29 | 3 |
| | 30-49 | 4 |
| | >=50 | 5 |
| **Nodule Count** | 0-9 | 1 |
| | 10-12 | 2 |
| | 20-29 | 3 |
| | 30-49 | 4 |
| | >=50 | 5 |
| **Spiculation** | Presence | 1.0 |
| | Absence | 0.0 |
| **Emphysema** | Presence | 1.0 |
| | Absence | 0.0 |
| **Upper lung nodule** | Presence | 1.0 |

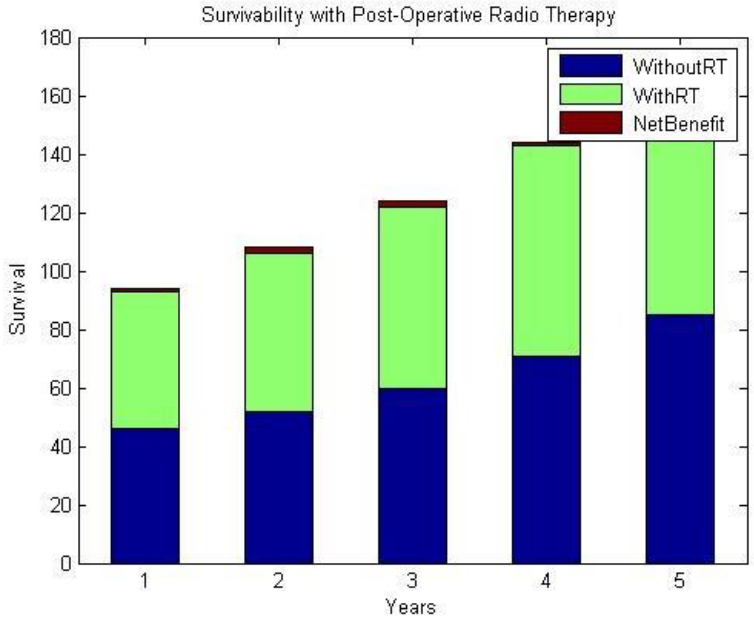
| | | |
|---|---|---|
| | Absence | 0.0 |

*4.4 TNM Stage and Cancer Stage*

The stage of Non-small cell lung cancer (NSCLC) can be predicted with the most common staging system, the TNM System developed by the International Union against Cancer. It is used only in NSCLC staging, when surgery is considered. We use different staging for both non-small cell lung cancer and small cell lung cancer.

*4.5 Recurrence rate with Post-Operative Radio Therapy*

The survival benefit of post-operative radiotherapy (PORT) for non-small cell lung cancer (NSCLC) stays notorious.




So, we construct a survival prediction model to enable individualized predictions of the net survival benefit of PORT for NSCLC cancer patients based on the tumour characteristics such as Histology, Tumour extent and node extent.

## V.DISCUSSION AND CONCLUSION

The earlier diagnosis of Lung cancer is a key to effective treatment. In this paper a novel prognosis method in the combination of regression and random forest classifier technique was used to build a lung cancer predictive model. Cancer associated death is increasing intensely. This rate can be reduced only with earlier prediction. But the worst case is that most people avoid cancer screening owing to cost and time associated with it. The model reduces the cost for different medical tests and helps the patients to take precautionary measures well in advance. It leverages the power of both clinical and genomic data to prefigure the cancer risk. It also describes the functional assessment stage of a cancer affected patient. In future this prognosis model can be designed as a web based application and can be implemented in remote areas, to imitate the human diagnostic expertise for predicting the disease. A more efficient model can be built by using different techniques and algorithms. Similar model is to be built for other types of cancers.

## ACKNOWLEDMENT

I would like to take this opportunity to express my profound gratitude and deep regard to my guide, Prof. M. Revathi, CSE, Kingston Engineering College, for her exemplary guidance, valuable feedback and constant encouragement in completing this paper. Her valuable suggestions were of immense help in getting this work done. Working under her, was an extremely knowledgeable experience. Also, I would like to extend my sincere gratitude to my parents and my sister for their constant support and encouragement throughout my education.

## REFERENCES

[1] Tasnuba Jesmin, Kaswar Ahmed, Md. Badrul Alam Miah (2013) Brain Cancer Risk Prediction System Using Data mining. International Journal of Computer Applications, Volume 61- No.

[2] Seyyid Ahmed Medjahed, Tamazouzt Ait Saadi, Abdelkader Benyettou (2013) Breast Cancer Diagnosis using K-Nearest Neibhor with Different Distances And Classification Rules. International Journal of Computer Applications, Volume 62- N0.1.

[3] Kawasar Ahmed, Tanuba Jesmin, Md.Zamilur Rahman (2013)Early Prevention and Detection of Skin Cancer using Data mining. International Journal of Computer Application, Volume 62-No.4.

[4] V.Kroshnaiah, Dr.G.Narsimha, Dr.N.Subhash Chandra (2013) Diagnosis of Lung Cancer Prediction System Using Data mining Classification Techniques International Journal Of Computer Science And Information Technologies, Vol 4(1), 39-45.

[5] Ada Ranjneet Kaur (2013) A Study of Lung Cancer Using Data mining Classification Techniques. International Journal of Advanced Research in Computer Science and Software Engineering, Vol 3, Issue 3.

[6] Ferlay J et al GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC Cancer Base No. 11 Lyon, France: International Agency for Research on Cancer; 2013. website http://globocan.iarc.fr, accessed 1 June 2015.

[7] Mittra, I. et al. A cluster randomized, controlled trial of breast and cervix cancer screening in Mumbai, India: methodology and interim results after three rounds of screening. Int. J. Cancer 126, 976–84 (2010).

[8] Parkin, D. M. & Boyd, L. P1-277 The fraction of cancer attributable to lifestyle and environmental factors in the UK in 2010. J. Epidemiol. Community Heal (2011).

[9] Nagarajan S, Chandrasekaran RM. Design and implementation of expert clinical system for diagnosing diabetes using datamining techniques. IJST. 2015 Apr; 8:771–6.

[10] Kalaiselvi C Nasira GM. Classification and Prediction of heart disease from diabetes patients using hybrid particle swarm optimization and library support vector machine algorithm. International Journal of Computing Algorithm (IJCOA). 2015 Mar; 4:1403–7.

[11] N.Revathy, Dr.R.Amalraj(2011) Accurate Cancer Classification using Expression of Very few Genes. International Journal of Computer Application Volume 14 – No.4.

[12] K.Rama Lakshmi and S.Prem Kumar (2013) Utilization of Data mining Techniques for Prediction and Diagnosis of Major Life Threatening Diseases Survivability- Review International Journal of Scientific & Engineering Research, Vol 4, Issue 6.

[13] Kalaiselvi C, Nasira GM. A new approach for the diagnosis of diabetes and cancer using ANFIS. World Congress on computing and communication technologies, WCCCT-IEEE conference; 2014 Feb. p. 188–90.

[14] Calculate your risk.Australian Government [Oneline]. Available: http://canceraustralia.nbocc.org.au/risk/caculator.php 110-115.

[15] Alessandra Saldanha de. Impact of Diabetes on Cardiovascular Disease: An Update. International Journal of Hypertension. 2013; 653789.

[16] S. Dey, R. Gupta, M. Steinbach, V. Kumar, Integration of Clinical and Genomic data: a Methodological Survey, Technical Report, Department of Computer Science and Engineering University of Minnesota, 2013

[17] M. H. van Vliet, H. M. Horlings,M. J. van de Vijver, M. J. Reinders, L. F. Wessels, Integration of Clinical and Gene Expression Data Has a Synergetic Effect on Predicting Breast Cancer Outcome, PLoS ONE, 7(7), 2012

[18] S. Watt, W. Jiao, A. M. Brown, T. Petrocelli, B. Tran, T. Zhang, J.D. McPherson, S. Kamel- Reid, P.L. Bedard, N. Onetto, T.J. Hudson, J. Dancey, L.L. Siu, L. Stein, V. Ferretti, Clinical genomics information management software linking cancer genome sequence and clinical decisions, Genomics, 102(3), 140-147, 2013.

[19] M. Kanehisa, S. Goto, Y. Sato, M. Furumichi, M. Tanabe, KEGG for integration and interpretation of large-scale molecular data sets, Nucleic Acids Research 40(D1), D109-D114, 2012.