# SURVEYON DATA MINING CLASSIFICATION TECHNIQUES FOR PREDICTION OF LUNG CANCER

Priyanka Rajpoot[1], Mahesh Parmar[2]

**Abstract—Lung cancer is the world's leading cause of deathfrom cancer. Medical field has become one of the researchers ' favorite topics in the current epoch of technology, and cancer is one of them. Healthcare researchers and clinicians on software advances To Effectively managing data and make inferences from large and heterogeneous data on healthcare. Cancer has different symptoms like a lump, excessive bleeding, weight loss, etc. There is no need for cancer of all kinds of tumors. We addressed various data mining methods used for breast cancer, lung cancer, and blood cancer in this surveypaper. They concentrate on current studies using the data mining method to improve the risk factors for breast, lung, blood cancer are diagnosis and prognosis.**
**Keywords—Data Mining,Lung Cancer, Health care needs, Cancer Diseases, Prediction techniques.**

## 1. INTRODUCTION

Data mining (DM) is hidden predictive data extraction Relationships and unknown data, trends, and observations by analyzing large data sets using traditional, hard-to-find and detect statistical methods. For decision-making, Using DM tools. Techniques of Prediction and classification are used to predict the unknown values of the template generated by the classification technique. A variety of DM Data associations and regularities can be established using techniques, to obtain legal information and to predict the value of dependent variables: Support Artificial Neural Network (ANN),Bagging Algorithm, Vector Machine (SVM), Neighborhood Nearest (KNN), Decision Tree, Naive Bayed, Bagging Algorithm, Naive Bayed, etc. DM is an important step in the analysis of database knowledge (KDD), Data cleaning, data integration, data collection, pattern recognition,& data mining knowledge identification iterative process.DM was widely used in interaction, stock market prediction, marketing, finance, education, credit evaluation, risk prediction, data acquisition,scientific discovery, health and medicine, marketing, identification of fraud, stock market forecasting, health and medicine, etcHowever, DM has a significant presence in all medical fields for the diagnosis of diabetes, lung cancer, breast cancer, heart disease, kidney failure, hepatitis, diabetes, lung cancer, etc. DM applications include data Improved health analysis, prevention of different hospital errors, early detection, prevention of various hospital errors, policy-making and prevention of various diseases through false insurance claims, value Cost savings and saving more lives by cutting the death rate for more money. Automated medical diagnosis allows doctors with less time to determine the right disease[1].

Lung cancer among different types of cancer worldwide is one of the most malignant tumors. Every year, roughly 1.2 Lung cancer has been diagnosed in millions of people and nearly 1,1 million people have been diagnosed. it. Much effort has been made to clarify the treatment and prognosis of lung cancer pathology. Histopathology, for example, the method of CT scanning, is widely used for the diagnosis of lung cancer and yet these Histopathology-based approaches are unsuccessful in predicting disease progression & clinical outcome Because of the nuanced nature of cancer and the human subjectivity. When cancer is detected & identified with microarray technology, gene expression profiles have been shown to be more active than histopathology [2]. The main reason for this is that only 15% of the early diagnosis of lung cancerwhich indicates that early diagnosis is key to improving the lung cancer prognosis. Medical imaging scans, minimally invasive interventional testing, tumor marker identification, and currently sputum cytology analysis are the main methods of a lung cancer diagnosis. However, these procedures have limits on Low sensitivity, time-consuming, invasive and other factors make research into an easy & rapid approach to early analysis of lung cancer key to improving survival of patients with lung cancer[3].

## 2. HEALTHCARE IN DATA MINING

It is difficult for A health care professional can identify high-risk COPD patients, but knowledge of the clinical history of the patient can make it easy. Data mining is the practice of using traditional statistical methods to analyze large data sets to find patterns, relationships, and details that are very difficult to identify. In new areas, DM in the healthcare industry places great emphasis on doctor's perspectives on how patients can recover from a disease, as well as a deeper understanding of medical data. Researchers used DM techniques in the medical diagnosis of several diseases, such as cancer, diabetes, heart disease, and stroke, as well as techniques such as neural networks, bagging algorithms, kernel density, naive Bayes, decision tree, Automatically specified groups and vector supporting machines with different levels of accuracy [4].

---

[1,2] Deprtment of computer science and engineering , Madhav institute of technology and science,gwalior

*2.1 The need for Data Mining In Healthcare*

This produces thousands of gigabytes of data every day in the form of statistical, item characteristics, patient Records of health. Such raw data are unstructured and disseminated widely through various medicine and pharmacy sections [1]. Such data must be evaluated in order to provide a systematic result and insight into how it can be useful for companies to compete in the market.It falls into drugs, disorders, patient demography, and etc. information analytics. Data analysis is done by DM, using criteria of existing drug components, emerging patterns, based on previous medical research awareness or information [5]. Although the efficiency and scalability of DM algorithms need to be improved, at least current It is necessary to improve data mining algorithms, at least current DM algorithms are better able to work large data sets than conventional analyzes of data. Algorithms are better than conventional data analysis capable of working with large datasets. DM produces results that have great value to their research, which can turn this business into a turning point and lend itself to changing fortunes.Because predefined trends, models, relationships can be identified, data mining can be used to predict and decide [6 ].

## 3. CANCER DISEASES PREDICTION

Cancer is the deadly disease caused by changes in the body's normal cells, resulting in unchecked cell growth leading to tumor, with the exception of leukemia, which This is the primary cause of cancer. If you are not found in a tumor timely manner, it develops and spreads through the bloodstream into surrounding areas and affects the gastrointestinal, organ and circulatory system, causing severe health effects that are The biggest cause of death. Lung, breast, stomach, and liver cancer are more likely to affect men. Females are more likely to suffer from lung, cervix, breast, and cancer of the stomach. Cancer is also a set of related diseases. If the disease is not properly treated or diagnosed in time, this malignant disease can even cause death in most cases.

*3.1 Classification of Cancer*

Overall, over 200 cancer forms were found. The cancer name It is the type of cell that has formed them the name and classification of their epithelial or squalor cell. Certain types of cancer which starts in the following different categories of cells. For example, Lung cancer affects the lung cells, which is referred to as lung cancer, brain cancer, and other cancers. Often We were responsible for the type of cell that formed them their epithelial name and classification or squalor cell. Some forms of cancer beginning in specific cell types are as follows.

Sarcoma:Leiomyosarcoma, lip sarcoma, and osteosarcoma are a type of soft tissue or malignant bone tumor. Most of the sarcoma is bone cancer. Leiomyosarcoma, malignant fibrous histiocytoma, capos sarcoma, lip sarcoma, and protuberant sarcoma in sarcoma are the most common soft forms of tissue sarcoma.

Leukaemia:Is a tissue cancer that shapes the cells of the blood, White blood cells and bone marrow. There's a lot of subtypes, including chronic lymphocytic leukemia and lymphocytic leukemia. All four leukemia types are categorized based on how quickly cancer on this type of blood cell starts to get worse.

Carcinoma: The epithelial cells emerge. It may invade the tissues around it. Breast, prostate, lung, and colon are examples of cancer carcinoma forms. Carcinomas have a special name because they are the type of cancer in the epithelial cells containing fluids that are formed in specific epithelial cells such as Aden carcinoma.

Lymphoma: Cancer that affects the lymph system, spreads all over the body and any part of the body can be affected.

*3.2 Causes of Cancer*

Tobacco:tobacco had as many as 80 substances affected by cancer. Once nicotine is inhaled like a cigarette, the contaminants enter the lungs and are then transferred through the bloodstream and the skin.

Overweight or Obesity:People are more prone to cancer because of higher levels of insulin overweight. It also increases the chance of cancer in women with a food tube, breast, liver, gallbladder, and cancer of the uterus.

Carcinogens: Genes coded messages telling them which protein they're going to produce. Changes in genes such as harm may interrupt their normal behavior or function, and then cells are Cancer is more likely to be identified as carcinogenic.

Alcohol: Alcohol is one of the major causes of cancer. Nearly all alcoholic beverages can cause cancer. This induces a number of cancers, including cancers of the mouth, ear, liver, and breast.

Infection: The risk of cancer is mainly due to infection with the virus. due to various forms of infection[7 ].

The Immune System: The risk of cancer is higher for people with weak immune systems. Those with HIV and AIDS have weakened the immune system is more resistant to cancer.

## 4. PREDICTION TECHNIQUE

Each paper gives an overview of the different techniques used. predictive purposes in data mining.

*4.1 Association Rule Learning*

Law education in partnerships is a very different method ofdatabase patterns and finds specific database rules. (i.e. commonly occurring observations). Training the rule of association helps in the process of decision making. Therefore, in market basket research, internet application Mining, clinical Diagnosis of sets of X and Y of various items, the learning of association rules

is very useful. partnership rules use a minimum support value specified by the user and a minimum confidence value specified by the user.

### 4.2 Classification
Classification is a way to find rules that classify knowledge In various groups. The classification process makes the large database a set of similar observations. A classification implementing Algorithm It is known as a classifier, a mathematical method with anAlgorithm for classification. This technique's main task is to identify and organize the related observation from the large dataset into a package.

### 4.3 Cluster Analysis
Clustering and researching is a methodology identical to that characteristic to group the objects in the same group. It is a data grouping method bioinformatics, cluster analysis is widely used. Clustering is a segmentation-like technique that Similar results for groups. Using the principle of association, the observations produce different patterns And then group them according to that pattern by the clustering technique. The same type of data is processed in the same groups in this approach and the groups are called clusters. Artificial neural network and closest search for neighbors are various techniques used in cluster analysis. Clustering helps to analyze sequences, group items.

### 4.4 Decision Tree
A decision-making tree, tree-like statistical system. -Tree branch, if the condition is met, denotes the leaves and the state consequence. With a question or a condition, that division has been recorded for two or more responses. Each answer can lead to a different Decision tree classifies data by circumstance without missing information and assists in the decision-making process. The technique of supporting medical diagnosis, vector machines can also identify handwritten characters. Vector supporting systems are theoretically dependent on numerical learning and help minimize the classification error and minimize the risk.

### 4.5 Neural network
Also used is the neural network to identify and model different patterns. The neural network receives a set of inputs Then uses one or more inputs to estimate. The numerical efficiency of the neural network. Neural networks are useful in detecting fraud, predicting consumer response, photo interpretation, and much more. for outer analysis or clustering, neural networks also help.

### 4.6 Regression Analysis
Analysis of regression makes similar observations from the various trend observations set. It analyzes the relationship between two different variables and examines the value of the variable dependent on the value of the independent variable. change. This regulates the state of the independent variable. For forecasting, this approach is used. Regression analysis uses a formula known as the regression equation for the independent variable.

### 4.7 Sequential Pattern Mining
A sequence is a set of transactions which is ordered. Sequential mining activities as the name suggests the data sets an array of different patterns. The methodology identifies patterns that occur frequently. Sequential pattern mining helps to identify frequently occurring current trends or related events. Two types of sequential mining patterns and strings mining pieces array. The algorithm recognizes a number of common observations using different rules in the mining of objects. For business applications, sequential pattern mining is commonly used. Apriori algorithm is widely used in sequential pattern mining[ 6 ].

## 5. LITERATURE SURVEY
K. Roy et al.[2019] The purposeThis research aims to improve accuracy and establish specific value for early-stage identification The combination of biomedical image processing techniques and information knowledge exploration of lung carcinoma. The image of the lungs obtained from the scanning images of the CT(Computed Tomography) is preprocessed and segmentation is carried out in The Random Forest test was used by the Region of Interest (ROI) to distinguish distinct characteristics. The improvement in saliency has been performed and the use of SURF (Speed Up Robust Features) algorithms Based on Saliency Enhanced images using SVM, such as entropy, co-relationship, energy, and variance classification. Classification determines the healthy or unhealthy image(carcinoma) produced. Total goal vs. Mission calculation plot number is used to evaluate the system performance. Two algorithms were used to complete this entire phase-random forest algorithm and classification of SVM. Using SVM classification, the best results are obtained. This process is an effective maximum of 94.5 percent, 74.2 percent, 66.3 percent, and 77.6 %[7 ].
E. D'Arnese et.al.[2019] model to enable Radiologists and doctors to use non-small cell lung cancer (NSCLC) non-invasive in-vivo characterization to enhance the diagnostic process. In order to achieve an accurate segmentation, our pipeline uses the data from regularly collected PET and CT images: the exact textural characteristics are determined in the value volume observed (VOI) It provides information on the characterization of lung lesion by means of machine-learning algorithms. We

tested our pipeline on the actual data sets of a private hospital. For VOI segmentation, our method achieved a mean accuracy of 94.2±5.0 percent and demonstrated the ability of PET / CT characteristics to Differentiate Subtypes of initial and metastatic lesions or lung cancer [8].

D. S. Senthil and B. Aishwarya [2018]The proposed method with different lung cancer data will be implemented in Matlab. In addition, we will equate our proposed work with current Strategies and algorithms to prove that the research they suggest is the best. Our study project Focuses on whether the classification technique of lung cancer is normal or abnormal. Initially, the correct information will be extracted from the input data collection after pre-processing during the preprocessing phase; the resulting output will be fed to the choice of the function. The features were selected in this selection processUsing the firefly algorithm. The unique functionality of the support vector machine (SVM) is given classifier after selecting the method.; the information is classified as either ordinary or abnormal with the help of this classifier[9].

S. Bharati et.al.[2018] Focused on the detection of breast cancer with the aid of data mining techniques, using various classification algorithms. Intelligent data mining obtained from patients previously diagnosed with illness A new field of scientific innovation has been opened up. In this paper, naive Bayes ' classification capacity, To predict breast cancer, Logistic Regression, Multilayer Perception, Random Forest, and K-nearest neighbors UCI machine learning database data collection Breast Cancer Disease was observed [10].

A. Safiyari and R. Javidan[2017] The Dataset Analyzed surveillance, epidemiology and end results (SEER) various predictions of lung survival learning methods in this study. Prior to the implementation of classification models, data are preprocessed in several phases. Five common ensemble methods, Bagging, Dagging, Ada Boost and Random Sub Space, as well as Eight algorithms of classification, RIPPER, Stump decision, C4.5, SMO, Random Forest and Bayes Net, Logistic Regression, , Multi Boosting, Decision Stump, were assessed as the basic classification for survival of lung cancer.After treatment for 5 years, mortality risk was estimated. The qualities of Predictions are calculated in relationto accuracy & area under the ROC curve (AUC). Ada Boost Algo had the best efficiency compared to the other four approaches to the ensemble in increasing the performance of base classifiers. RIPPER's accuracy improved The decision stumps the reliability of the decision algo after 81.21% to 87.67%, & SMO algorithm from 88.88% to 88.98%. The value is 83.41% to 87.16%. So, too. that's right. Also, RIPPER AUC increased by 91,5%, up to 94,9% The decision-making AUC algorithm ranges from 81,2% to 93,9%, the J48 AUC algorithm ranges from 94% to 94,9% and the SMO AUC algorithm ranges from 50,0% to 92,1%. In contrast to other assembly approaches, the random subspace algorithm was the worst solution. The results showed empirically that ensemble methods were able to evaluate the importance of their specific classifiers and were acceptable methods for the cancer survival study [11 ].

Chauhan [2016] The proposed model is an efficient approach focused on classification in which principles of machine learning are used to identify Diseases for lung cancer. The algorithm has generated promising results, but it requires significant expertise in computing to be executed. In addition, it has been shown that certain comparison sets compare the proposed working model. Automated classification of diseases Using machine learning is often based on individual object segmentation features that can be difficult to automate[12].

Sadhu [2016] Experiments were performed on lung cancer data sets, Breast cancer, and acute myeloid leukemia(AML) standard sub-cluster mining algorithm. Results were found to match those reported in previous studies to a large extent. Few genes emerge from the respective diseases as indicative molecular biomarkers. The aim is to investigate potential molecular biomarkers of such diseases in order to detect them early & consistently. Can identify these diseases early and reliably, molecular biomarkers. A study with a methodology for common sub-cluster mining using FCM clustering was reported in this paper. Sub Cluster refers to peaks formed by overlapping clusters obtained separately from expression data from both normal and diseased samples[13 ].

A. Akayet al.[2015] Propose a two-step research process based On the user forum posts, defining user groups (Modules) and influential users to assess the user's output views on cancer treatment, positive and negative emotions, and on the side effects of care. To evaluate word frequency information from user forum posts, we used a self-organizing chart. We then introduced a new network-based modeling system for user interaction and used a network partitioning method to optimize reliability quality measurement. This allowed us to define customer feelings and use Word-frequency data information as well as network-based properties to recognize influential users within the modules that were collected. The methodology Could extend Work on smart mining of social media data to provide quick, up-to-date information on effectiveness public feedback on various treatments (or inefficiency) of potential treatments for the pharmaceutical industry, medical personnel, and hospitals, [14 ].

Liuet al.[2014] In this paper, data mining techniques and conventional statistical classification approaches were used to quantify & interpret SERS saliva array includes 59 Lung cancer patients & 18 healthy patients. The data was collected using the discriminating, Random Forests Algorithm (RF) and Fisher process, support vector Machine (SVM) and the models discussed lung cancer's auxiliary diagnostic effectiveness.SVM & RF diagnostic indices are lower than Fisher's discriminating test, and the best model of a lung cancer diagnosis can be believed to be determined. The results show that a new type of lung cancer diagnostic tool can be used to diagnose lung cancer on data mining through SERS.compared to healthy people [15.

Table 1: A comparative study of classification technique of lung cancer prediction

| Application | Advantage | Disadvantage | Goals |
|---|---|---|---|
| The SEER database is used to classify patients with lung cancer by means of supervised learning techniques, including linear regression, decimation tree, GBM, SVM, customized ensemble. Specialist training techniques are also applied to the SEER database.[16] | improve the accuracy of predicting cancer & helping to improve our basic understanding of cancer [25] | Numerous side effects, particularly including severe side effects, were more common under dacomitinib. Patient-reported symptoms worsened more frequently and also earlier [19]. | To inform decisions on patient care and that this dataset can be used to match the performance of these techniques with that of conventional methods[16 ]. |
| In order to take these approaches into account in dailies, it can improve our understanding of cancer development[19 ]. | the benefits of an integrated palliative care approach for lung cancer patients at the time of diagnosis and to analyze the main symptoms and its treatment.[24] | Deterioration of health-related quality of life was also more pronounced than under gefitinib[19]. | the goal of enabling predictive power correlation between different methods The prediction is regarded as an ongoing goal,[16 ]. |
| The use of ML methods will obviously improve cancer resistance, recurrence, and prediction of survival[19 ]. | The main benefits of this technique areto provide the earlier warning to the users [23] | The classification technique in the prediction of lung cancer is the problem of assigning a sample to a predefine class based on condition features.[21] | Classify lung cancer patients automatically into classes based on observable disease-specific variables to predict their survival[17 ]. |
| Analyzing and sensibly presenting information to provide useful information by using software tools and techniques. Predict the risk of lung cancer disease in patients[20 ]. | It is also cost & time-saving benefits to the user because it predicts not only lung cancer but also examining a different disease[23] | To physicians in the field of medicine, cancer remains a problem. Right now it is not possible to discover the actual cause and cure of cancer[9 ]. | Reduce the variability of lung cancer risk assessment and reporting between doctors who interpret it. [ 18 ] The report is in French only. |

## 6. CONCLUSION

In their survey papers, this paper addresses various use of data mining methods by researchers. The use of DM is discussed in a medical prediction area. The main focus is on using data mining to use different algorithms to predict cancer. A prototype system for the prediction of lung cancer is developed using techniques for the classification of data mining. The program derives hidden knowledge from a list of historical cancer of the lung. naive Bayes tend to be the most powerful model for predicting patients with lung cancer disease followed by IF-THEN law, Trees and Neural Network decision-making. The results of Decision Trees are easier to read and understand. only order to access detailed patient profiles, Decision Trees can be found on the drill. Naive Bayes was better off than trees because all the important medical predictors could be established.

## 7. REFERENCES

[1] Liu, W., Man, Z., Hua, L., Chen, A., Wang, Y., Qian, K., & Zhang, Y. (2014). Data mining methods of lung cancer diagnosis by saliva tests using surface-enhanced Raman spectroscopy. 2014 7th International Conference on Biomedical Engineering and Informatics. DOI:10.1109/bmei.2014.7002849

[2] Chauhan, D., &Jaiswal, V. (2016). An efficient data mining classification approach for detecting lung cancer disease. 2016 International Conference on Communication and Electronics Systems (ICCES).

[3] Zhang, Y. (2014). Data mining methods of lung cancer diagnosis by saliva tests using surface-enhanced Raman spectroscopy. 2014 7th International Conference on Biomedical Engineering and Informatics. DOI:10.1109/bmei.2014.7002849

[4] Gan, B., Zheng, C.-H., & Wang, H.-Q. (2014). A survey of pattern classification-based methods for predicting the survival time of lung cancer patients. 2014 IEEE International Conference on Bioinformatics and Biomedicine (Koppad, S. H., & Kumar, A. (2016). Application of big data analytics in the healthcare system to predict COPD. 2016 International Conference on Circuit, Power and Computing Technologies (ICCPCT). DOI:10.1109/iccpct.2016.7530248 BIBM). DOI:10.1109/bibm.2014.6999296

[5] Satyam Shukla+, DharmendraLal Gupta # and BakshiRohit Prasad*." Comparative Study of Recent Trends on Cancer Disease Prediction using Data Mining Techniques", International Journal of Database Theory and Application Vol.9, No.9 (2016), pp.107-118.

[6] Satyam Shukla+, DharmendraLal Gupta # and BakshiRohit Prasad*." Comparative Study of Recent Trends on Cancer Disease Prediction using Data Mining Techniques", International Journal of Database Theory and Application Vol.9, No.9 (2016), pp.107-118.

[7] K. Roy et al., "A Comparative Study of Lung Cancer detection using supervised neural network," 2019 International Conference on Opto-Electronics and Applied Optics (Optronix), Kolkata, India, 2019, pp. 1-5.

[8]   E. D'Arnese, G. W. di Donato, E. del Sozzo and M. D. Santambrogio, "Towards an Automatic Imaging Biopsy of Non-Small Cell Lung Cancer," 2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI), Chicago, IL, USA, 2019, pp. 1-4.

[9]   D. S. Senthil and B. Aishwarya, "Predicting Lung Cancer Using Datamining Techniques With the AID of SVM Classifier," 2018 Second International Conference on Green Computing and Internet of Things (ICGCIoT), Bangalore, India, 2018, pp. 210-216.

[10]  S. Bharati, M. A. Rahman and P. Podder, "Breast Cancer Prediction Applying Different Classification Algorithm with Comparative Analysis using WEKA," 2018 4th International Conference on Electrical Engineering and Information & Communication Technology (IEEE ICT), Dhaka, Bangladesh, 2018, pp. 581-584. DOI: 10.1109/CEEICT.2018.8628084

[11]  A. Safiyari and R. Javidan, "Predicting lung cancer survivability using ensemble learning methods," 2017 Intelligent Systems Conference (IntelliSys), London, 2017, pp. 684-688. DOI: 10.1109/IntelliSys.2017.8324368

[12]  Chauhan, D., &Jaiswal, V. (2016). An efficient data mining classification approach for detecting lung cancer disease. 2016 International Conference on Communication and Electronics Systems (ICCES).

[13]  Sadhu, A., & Bhattacharyya, B. (2016). Discovery of cancer linked biomarker genes through common subcluster mining. 2016 International Conference on Bioinformatics and Systems Biology (BSB). DOI:10.1109/bsb.2016.7552153

[14]  A. Akay, A. Dragomir and B. Erlandsson, "Network-Based Modeling and Intelligent Data Mining of Social Media for Improving Care," in the IEEE Journal of Biomedical and Health Informatics, vol. 19, no. 1, pp. 210-218, Jan. 2015.

[15]  Liu, W., Man, Z., Hua, L., Chen, A., Wang, Y., Qian, K., & Zhang, Y. (2014). Data mining methods of lung cancer diagnosis by saliva tests using surface-enhanced Raman spectroscopy. 2014 7th International Conference on Biomedical Engineering and Informatics. DOI:10.1109/bmei.2014.7002849 .

[16]  https://www.ncbi.nlm.nih.gov/pubmed/29132615

[17]  https://www.ncbi.nlm.nih.gov/pubmed/28910336

[18]  https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6037965/

[19]  Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V., & Fotiadis, D. I. (2015). Machine learning applications in cancer prognosis and prediction. Computational and Structural Biotechnology Journal, 13, 8–17. DOI:10.1016/j.csbj.2014.11.005

[20]  https://www.researchgate.net/publication/279517741_A_Comparative_Study_of_Data_Mining_Classification_Techniques_using_Lung_Cancer_Data

[21]  http://ijiset.com/vol3/v3s2/IJISET_V3_I2_07.pdf

[22]  Oh, J. H., Al-Lozi, R., &Naqa, I. E. (2009). Application of Machine Learning Techniques for the Prediction of Radiation Pneumonitis in Lung Cancer Patients. 2009 International Conference on Machine Learning and Applications. DOI:10.1109/icmla.2009.118

[23]  https://www.ncbi.nlm.nih.gov/pubmed/26655954

[24]  https://www.irjet.net/archives/V2/i8/IRJET-V2I8175.pdf.

[25]  https://www.scienceopen.com/document_file/d20a673a-f4f3-4cab-a63a-4c7cbfa7d4c9/PubMedCentral/d20a673a-f4f3-4cab-a63a-4c7cbfa7d4c9.pdf