

A Survey on Diagnosis of Lung Cancer Diseases Using Machine Learning Approaches

Rekha Awasthi^{1*}, Vaibhav Chandrakar², Vijayant Verma³, Poonam Gupta⁴

^{1,3}Disha College, Raipur, India

^{2,4}CCEM, Raipur, India

Corresponding Author: awasthirekha151@gmail.com

DOI: <https://doi.org/10.26438/ijcse/v7i3.371374> | Available online at: www.ijcseonline.org

Accepted: 21/Mar/2019, Published: 31/Mar/2019

Abstract— In the field of Healthcare, cancer finding is the testing issues and furthermore a considerable lot of the exploration has centered to enhance the performance to get satisfactory outcomes in the specific territory. To analyze a Lung cancer is a troublesome errand in medical research. To beat this testing errand, the numerous analysts use data mining methods were connected to predict the many kind of disease. In this examination we studied and make compression of various classifications to classify and predict the lung cancer illness.

Keywords— Lung Cancer Detection, Segmentation, Feature Extraction and Classification.

I. INTRODUCTION

The tumor which begins from the lungs is known as lung cancer. Lung cancer is a real general health issue wherever the world. Lung cancer is the uncontrolled development of abnormal cells that begin off in one or the both lungs [1]. Lung cancer is the second most ordinary tumor in the two men and women and is primary wellspring of cancer death among the two men and women. Most lung cancers could be avoided, in light of the fact that they are identified with smoking or less regularly to introduction to radon or environmental factors [2]. Lung cancer can be grouped into two primary sorts: Small Cell Lung Cancer (SCLC) and Non-Small Cell Lung Cancer (NSCLC). Small Cell Lung Cancer represents 20% of lung cancers, while Non-Small Cell Lung cancer represents the staying 80% [2]. As per the World Health Organization (WHO), 7.6 million deaths globally every year are brought about by cancer; cancer speaks to 13% of every global death. As observed underneath, lung cancer is by a wide margin the main cancer killer [1]. As lung development cells spread and use a more prominent measure of the body's vitality, it is conceivable to exhibit manifestations that may likewise be related with numerous different ailments [1]. These include:

- Fever
- Fatigue
- Unexplained weight loss
- Pain in joints or bones
- Problems with brain function and memory
- General weakness

- Bleeding and blood clots

Beginning time of lung cancer is hard to identify at stage I and II. The vast majority with NSCL are analyzed just at just at stage III and IV.

Types of tumors:

- Benign
- Malignant

A. Benign

If the tumor is benevolent, at that point the proportion of the tumor is less than 3mm. This is starting level of infection tumor. Under this class is easily reparable.

B. Malignant

In case the tumor is unsafe, at that point the range of the tumor is more unmistakable than 3mm. This is a wild level of development tumor. Under this class isn't reparable.

In medical Imaging utilized diverse kinds of pictures are X-beam, Computed Tomography (CT), Magnetic Resonance Imaging (MRI), and so on. Be that as it may, for the recognition of lung finding CT pictures are favored in view of good clearness, low commotion and exceptionally simple to ascertain the mean and change of CT scan pictures. CT scan pictures are immediately acquired and don't hurt the bones of the patient. CT is the most much of the time utilized imaging system in the determination of lung cancer [3]. CT pictures give 3D analysis of the inward body parts and organ analysis is simple since it is taken at various points. The identification procedure separated into four sections: Lunge

segmentation, Image Enhancement, Classification and Feature Extraction.

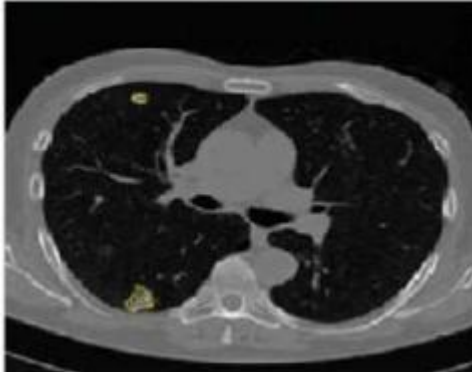


Fig.1: A sample lung image [4]

II. LUNG CANCER

In medical image processing, the location of different sorts of cancer influencing diverse organs is an essential continuous research territory. Lung cancer is a standout amongst the most regular cancers worldwide. Like bosom cancer in female, lung cancer is the main cancer site in males, containing 17% of the complete new cancer cases and 23% of the absolute cancer deaths. Two noteworthy kinds of Lung cancers are adenocarcinoma (counting bronchi alveolar carcinoma) speaking to about 40% and squamous cell carcinomas speaking to around 25-30%. Current examinations concerning early discovery and adjuvant chemotherapy intensely depend on the best possible organizing of patient cancer type. Before PC helped classification, the cell segmentation is an essential. Cancer location ought to be increasingly exact since its reports will be utilized for determination. Such huge numbers of specialists have given diverse methods. Biopsy is a best quality level for cancer identification including lung cancer. The utilization of fine needles or biopsy equipment to the influenced region is progressively painful to the patient. Normally, surgery is required for the analysis of lung cancer which will be costly and painful. Another usually utilized technique for lung cancer recognition is the analysis of CT pictures. However, the information acquired from CT has the issue of force variety. Henceforth, there is opportunity to get blame outcomes and it might prompt wrong finding and other related issues, which hurt medical practices. So there is a need of a substitute methodology for lung cancer recognition and it drove them to the analysis of cell pictures, which is superior to anything the CT image analysis. For this analysis they utilized the sputum of the influenced individual which can be acquired without surgery or some other destructive strategy.

III. LITERATURE SURVEY

A. Agrawal et al.[5], the ensemble voting of five decision tree based classifiers and meta-classifiers was found to result in the best expectation performance as far as accuracy and territory under the ROC curve. Further, we have built up an on-line lung cancer result calculator for assessing risk of mortality following 6 months, 9 months, 1 year, 2 year, and 5 years of conclusion, for which a smaller non-redundant subset of 13 attributes was cautiously chosen utilizing attribute selection techniques, while attempting to hold the prescient power of the original set of characteristics.

Bhola A.et al.[6], Author present a review of different cancer classification techniques and assess these proposed strategies based on their classification accuracy, computational time and capacity to uncover gene information. Creator have additionally assessed and presented different proposed gene selection strategy. In this paper, a few issues identified with cancer classification have additionally been examined.

V. Kirubhaet al. [7], This paper utilizes 'Tobacco use' data, so as to classify the risk factor level of Lung Cancer based on the tobacco use percentage of individuals. It likewise looks at data mining classification algorithms, for example, Naïve Bayes, Random Forest, Random Tree and REP Tree for performance analysis. The REP tree algorithm furnishes better outcomes when contrasted and different algorithms.

K. Jayasurya et al. [8], BN model structure and parameter learning distinguished gross tumor volume size, performance status, and number of positive lymph nodes on a PET as prognostic factors for two-year survival. At the point when approved in the full validation set of Ghent, Leuven, and Toronto, the BN show had an AUC of 0.77, 0.72, and 0.70, individually. A SVM demonstrate based on similar factors had a general more regrettable performance AUC 0.71, 0.68, and 0.69 particularly in the Ghent set, which had the most elevated percentage of missing the essential GTV size data.

Adaet al. [9], In this paper Histogram Equalization is utilized for preprocessing of the images and highlight extraction process and neural system classifier to check the condition of a patient in its beginning period whether it is normal or abnormal. After that we anticipate the survival rate of a patient by extracted features. Experimental analysis is made with dataset to assess the performance of the diverse classifiers. The performance is based on the right and wrong classification of the classifier. All experiments are directed in WEKA data mining device.

A.Priyanga et al. [10], Author has proposed the cancer prediction system based on data mining. This framework gauges the risk of the bosom, skin, and lung cancers. This framework is approved by contrasting its anticipated outcomes and patient's earlier medical information and it was examined by utilizing weka framework. The primary point of

this model is to give the prior warning to the users, and it is additionally cost efficient to the user.

Thangarajuet al. [11], Lung cancer is the main source of cancer death in the United States, among the two men and women. The two fundamental sorts are little cell lung cancer and non-little cell lung cancer. These sorts are analyzed based on how the cells look under a microscope. Individuals

who smoke have the most serious risk of lung cancer. The risk of lung cancer increments with the time span and number of cigarettes they have smoked. On the off chance that they quit smoking, even subsequent to smoking for a long time, they can fundamentally decrease his/her chances of creating lung cancer. In this work, author applies classification procedures on a dataset of lung cancer patients based on smoking and non-smoking individuals.

TABLE I. Comparisons of various techniques and method used in existing system

Ref. No.	Datasets	Classification Algorithms	No. of Instances	Accuracy	Limitations
5	SEER data	J48	5754 Instances	91.4%	Need to deal with unbalanced data.
6	Leukemia dataset	KNN	34Instances	84.6%	Need to develop more systematic and productive classification algorithms.
7	“Tobacco use” data.	REP Tree	3777Instances	79.58%	Need to improve the performance of applied algorithms and reduce the execution time by applying the data mining techniques.
8	Leuven and Ghent data sets	Bayes Network	322Instances	77%	Need to represent and integrate different aspects and stages of a disease in one unified framework
9	909 Lung CT-Scan images dataset	Neural Network	909Instances	96.04%	Need to overcome with the issues of increased size of dataset such as uploading data, managing feature set, increased execution time of classification algorithms etc. can be considered
10	Lung CT-scan data set	ID3	463Instances	100%	Need to improve the accuracy
11	UCIrepository dataset	Naïve Bayes	303Instances	83.4%	Need to design the clinical decision support system for mining Lung Cancer.
5	SEER data	SVM	34 Instances	94%	Need to develop on-line cancer outcome calculators for patients.

IV. CONCLUSION

Lung cancer is one of the real reasons for death in men and women. In Lung Cancer, look into eventually to enhance the nature of Healthcare and Lung Cancer patients. Now and again even in the propelled level Lung cancer patients does not demonstrate the manifestations related with the Lung cancer. There are numerous patients did not know Lung cancer illness in beginning time, on the grounds that the absence of mindfulness. In this way, the avoidance of lung cancer is required in diminishing life losses. By this experiment results we can unmistakably predict the lung cancer infection, which can be utilized to caution the general population before to keep away from the undesirable drinking habits, smoking, and intake of contaminated food, obesity and so forth, for keep from the lung cancer illness. In this investigation of classification Techniques utilized in a few dataset is help to enhance absence of mindfulness in the Lung Cancer patients.

REFERENCES

- [1] <http://www.medicalnewstoday.com/info/lung-cancer/> time: 1:33pm date: 7/10/2015
- [2] "Lung Cancer Prevention and Early Detection", Available at: <http://www.cancer.org/acs/groups/cid/documents/webcontent/acspc-039558-pdf.pdf>
- [3] Dandil E, Cakiroglu M, Eksi Z, Ozkan M, Kurt O.K, Canan A, "Artificial Neural Network-Based Classification System for Lung Nodules on Computed Tomography Scans", Soft Computing and Pattern Recognition (SoCPaR), 2014 6th International Conference, Aug 11-14, ©IEEE2014, DOI:10.1109/SOCPAR.2014.7008037, pp. 382-386.
- [4] Smitha P et al., "A review of medical image classification techniques, "International conference on VLSI, communication & instrumentation (ICVCI) 2011 proceedings published by International journal of computer applications, 34-38 pages.
- [5] Ankit Agrawal, Sanchit Misra, En. At Al, A Lung Cancer Outcome Calculator Using Ensemble Data Mining On Seer Data, Biokdd 2011, August 2011, San Diego, Ca, Usa.
- [6] Amit Bhola, Machine Learning Based Approaches For Cancer Classification Using Gene Expression Data, Machine Learning And Applications: An International Journal (Mlajj) Vol.2, December 2015
- [7] V. Kirubha, Comparison Of Classification Algorithms In Lung Cancer Risk Factor Analysis, International Journal Of Science And Research (Ijsr) Volume 6 Issue 2, February 2017.
- [8] K. Jayasurya, G. Fung, S. Yu, C. Dehing-Oberije, D. De Ruyscher, A. Hope, W. De Neve, Y. Lievens, P. Lambin, A. L. A. J. Dekker, Comparison Of Bayesian Network And Support Vector Machine Models For Two-Year Survival Prediction In Lung Cancer Patients Treated With Radiotherapy, The International Journal Of Medical Physics And Research, Vol. 37, No, 4, (2010).
- [9] Ada, Early Detection And Prediction Of Lung Cancer Survival Using Neural Network Classifier, International Journal Of Application Of Innovation In Engineering Of Management (Ijaiem), Volume 2, Issue 6, June 2013
- [10] A. Priyanga, S. Prakasam, Ph.D, Effectiveness Of Data Mining – Based Cancer Prediction System (Dmbcps), International Journal Of Computer Applications, Vol. 83 – No 10, December (2013), Pp. 0975 – 8887.
- [11] Thangaraju, Barkavi, Karthikeyan, 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, No. 7, July (2014).