International Journal of Engineering and Technical Research (IJETR)
ISSN: 2321-0869 (O) 2454-4698 (P), Volume-13, Issue-2, July-December 2023

Prognosis of Lung carcinoma using Data Mining Techniques

Shreevalli P

***Abstract*— Lung carcinoma staging is an integral part of knowing what treatment choices are. Generally, cancers that are detected at the primitive stage are easier to treat but mostly people are suitable to live a long time with advanced stage compliant. The stage of lung carcinoma is found by combination of TNM Tumor size, Regional lymph node involvement and Metastasis status. Lung cancer is divided based on the cell of origin, Non small cell lung cancer stages range from one to four, Small cell lung cancer is described using two stages, limited and extensive. Data mining techniques have played vital role in transforming decision making process in medical field. Using mining techniques it is possible to extract patterns from datasets in determining carcinoma .Various techniques like Decision tree, k nearest algorithm, Support vector machines are used.**

***Index Terms*— SVM,KNN, Random forest, Non small cell and Small cell lung cancer**

I. INTRODUCTION

The significant global concern is cancer.Reduced access to care because of health care setting closures and fear of COVID-19 exposure rebounded in detainments in opinions and treatment that may lead to short term drop in cancer prevalence followed by supplement in advanced stage complaint and eventually increased mortality.

When we suppose our health, we might not incontinently suppose of our lungs, still our lungs are like any other part of our body in that they gradationally age and wear out .The average person who lives 60 can have taken 504 million breaths and will have lungs that look veritably different to that of 16 years old. Besides natural decline as we progress our lung health can also be defected by environmental changes, improper lifestyles and nasty practices.

Data mining techniques can link various clinical cancer related attributes of patients to their survival outcomes. It has been observed that medical experts researchers are interested in using mining tools for its high performance.

II. RELATED WORK

Enormous medical researchers are interested towards prediction of the disease, which helps to save lives.
It is very important to identify cancerous cells in advance and to take the necessary preventive measures in the initial stages

Gibbons et al. (2019) used supervised learning such as linear regression model, support vector machine, ANN etc. and predicted that SVM results an better accuracy of 96% as compared to other methods [7].

The Patient EXperience of Bodily Changes for Lung Cancer Investigation (PEX-LC) study has published a model for predicting lung cancer based on reported symptoms and signs among patients having undergone PHC investigation [6].

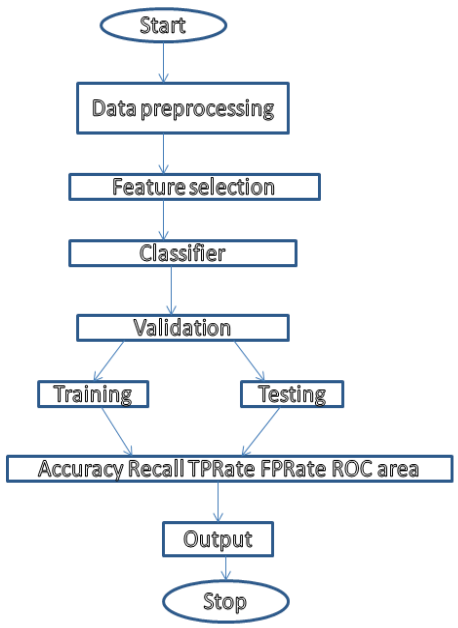
In [1] it is found that Relief and random forest generating best prediction.

In [2] it is evident that Lung cancer patients who get radiotherapy as a major aspect of their treatment are exposed to danger radiation- actuated lung in-jury known as radiation pneumonitis (RP). RP is a possibly deadly symptom leading to risky treatment. Later new ways re anticipated to manage medical experts to recommend concentrate on treatment dimension to patients at high peril of RP[3]

The contribution of calculations is assessed related to many element choice system and effect of the element determination on execution is future assessed[4]
Lung cancer diagnosis via symptoms and signs are sometimes downgraded in importance compared with screening. However, screening program for lung cancer have mostly been targeting high-risk smoking individuals [5], leaving the increasing group of never smokers without structured guidelines for early detection

III. TECHNIQUES USED

System Diagram



A .Decision tree

The decision tree for prognosticating the class of the given dataset, algorithm starts from the root of tree. Compares values of root with record(actual dataset)follows the branch and jumps to the next node

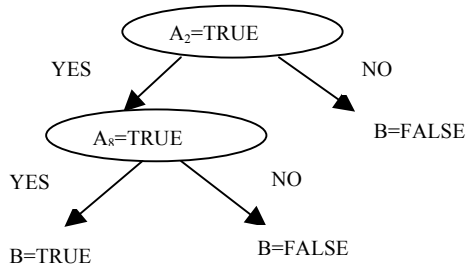


Fig 1.Sample Decision tree

B . Support Vector Machine

Support Vector Machine is a supervised learning algorithm that uses the Classification method to analyze data and predicate patterns.

SVMs are different from other classification algorithms because of the way they choose the decision boundary that maximizes the distance from the nearest data points of all the classes.

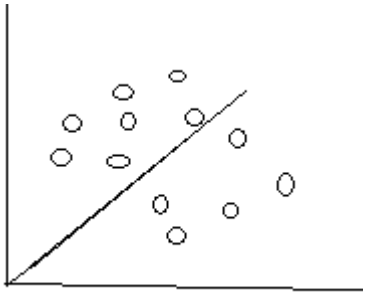


Fig 2 .SVM Classifier splits the data into two classes

C . K-Nearest Neighbor Classifier

The KNN algorithm is a supervised classification method. It's a simple algorithm that looks for the nearest fit.

K-nearest neighbors (KNN) algorithm uses 'feature similarity' to predict the values of new datapoints which further means that the new data point will be assigned a value based on how closely it matches the points in the training set

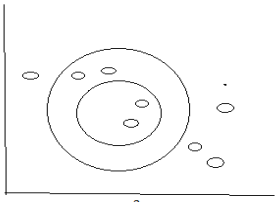


Fig 3 KNN Representation

D. Random forest

A random forest is a group of decision trees, A random forest will randomly choose features and make observations, build a forest of decision trees, and then average out the results.

Feature bagging also makes the random forest classifier an effective tool for estimating missing values as it maintains accuracy when a portion of the data is missing.

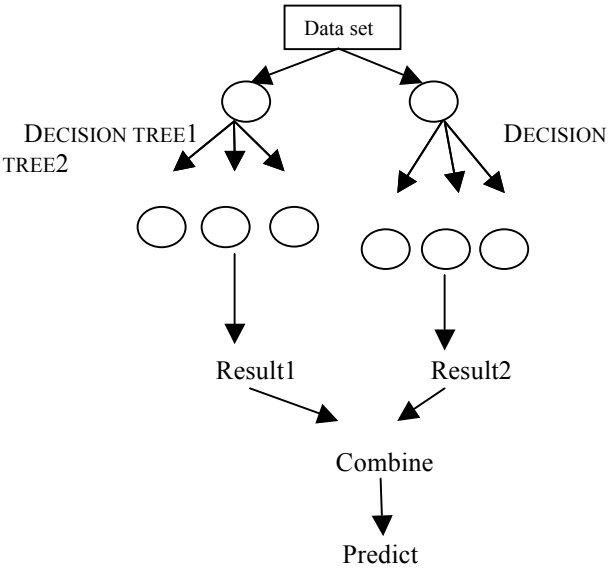


Fig 4.Representation of Random forest

E. Compare different techniques

	Logistic Regression	CART	Random Forest	KNN
1. Ease to interpret output	2	3	1	3
2. Calculation time	3	2	1	3
3. Predictive Power	2	2	3	2

IV EXPERIMENTAL RESULTS

Methodology	Training set	Testing set
K NN	75	37
Decision tree	10	12
SVM	0	38
Random forest	83	25

Different techniques results in tabular representation

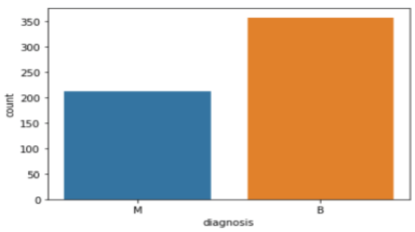


Fig 5 addresses the result of data analysis of cancer data which the number of have the disease and the number of didn't have the malignant growth.

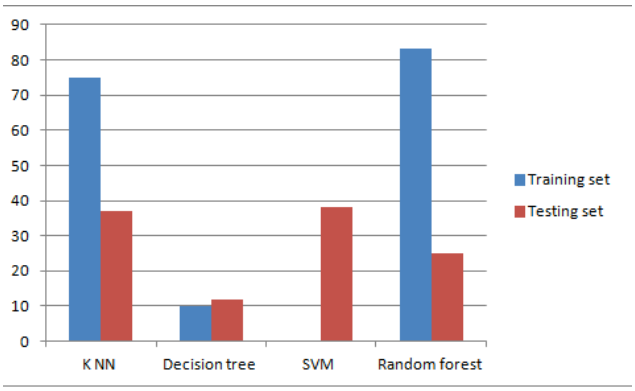


Fig. 6 Address the accuracy of each technique in graphical representation

V CONCLUSION

By acquiring the results of two data sets with the help of ML techniques, It is clear that the rate of prediction performed is better in KNN compared with decision tree, random forest, and SVM .To conclude that accuracy depends on the training data set methods before prior to acquainting them with testing.

VI . REFERENCES

[1] Small-Cell Lung Cancer Detection Using a Supervised Machine Learning Algorithm", march 2017.
[2] K.Srinivas,Dr.Mohammed Ismail.B "Testcase Prioritization With Special Emphasis On Automation Testing Using Hybrid Framework" Journal of Theoretical and Applied Information Technology Vol. 96. No 13 4180-4190 July 2018
[3] JaneeAlam, Sabrina Alam, AlamgirHossan, "Multi-Stage Lung CancerDetection and Prediction Using Multi- class SVM Classifier", 2018.
[4] Sarah Soltaninejad, Mohsen Keshani, FarshadTajeripour, "Lung Nodule Detection by KNN Classifier and Active Contour Modelling and 3D Visualization", April 2012.
[5] de Koning HJ, van der Aalst CM, de Jong PA, Scholten ET, Nackaerts K, Heuvelmans MA, et al. Reduced Lung-Cancer Mortality with Volume CT Screening in a Randomized Trial. N Engl J Med. 2020. Epub 2020/01/30. pmid:31995683
[6] Levitsky A, Pernemalm M, Bernhardson BM, Forshed J, Kölbeck K, Olin M, et al. Early symptoms and sensations as predictors of lung cancer: a machine learning multivariate model. Sci Rep. 2019;9(1):16504. Epub 2019/11/13. pmid:31712735; PubMed Central PMCID: PMC6848139.
[7] Sidey-Gibbons, J.A., Sidey-Gibbons, C.J.: Machine learning in medicine: a practical introduction. BMC Med. Res. Methodol. **19**(1), 64 (2019)
[8] Pellakuri Vidyullatha, Rajeswara Rao D, "Training and development ofartificial neural network models: Single layer feedforward and multi layer feedforward neural network",journal of Theoretical and Applied Information Technology Volume 84, Issue 2, 20 February 2016, Pages 150-156
[9] Mareedu Lakshmi Vihari, K Amarendra, Navvrula Anusha Recognition of Zeroday Exploit, International Journal of Engineering & Advanced Technology (IJEAT), ISSN: 2249- 8958, Volume: 08, Issue: 04, pp.1875-1877, April (2019).
[10] Xueyan Mei, Predicting Five-year Overall Survival in Patients with Non-small Cell Lung Cancer by ReliefF Algorithm and Random Forests, Feb 2017.
[11] Mohammad Ismail K.Naga Lakshmi, Y. Kishore Reddy, M. Kireeti, T.Swathi" Design and Implementation of Student Chat Bot using AIML and LSA" International Journal of Innovative Technology and Exploring Engineering (IJITEE) Volume-8 Issue-6,
[12] Mohammed Ismail B , K. BhanuPrakash, M. NagabhushanaRao" Collaborative Filtering-Based Recommendation of Online Social Voting" International journal of Engineering and Technology " Volume 7 issue 3 1504-1507 July 2018
[13]MohammadIsmail,V.HarshaVardhan,V.AdityaMounika, K. SuryaPadmini "An Effective Heart Disease Prediction Method Using Artificial Neural Network "International Journal of Innovative Technology and Exploring Engineering' at Volume-8 Issue-8, pp 1529-1532 June 2019.

Shreevalli P who is working as an Assistant Professor in Sambhram Academy of Management Studies and having 7 years of experience in teaching both BCA and MCA students. Her areas of interest are Networking, Data mining, Web programming, Image processing, Machine learning.