Prediction of Lung Cancer Using Machine Learning Algorithms

Ayush Mishra
School of Computer Science and Engineering
Vellore Institute of Technology Chennai
ayush.mishra2022a@vitstudent.ac.in

Dr.Monica KM
School of Computer Science and Engineering
Vellore Institute of Technology Chennai
monica.km@vit.ac.in

Abstract-The lungs are affected by the cell imbalance brought on by lung cancer. Early detection of lung cancer is crucial especially in countries with high levels of poverty and dense population. The frequency of cancer cases is increasing every day. Our health care system would benefit greatly from the early prediction of lung cancer. Therefore, most widely used technologies today is the application of machine learning. Machine learning is used to develop various algorithms. Machine learning algorithms are being created, and they will play a major part in having many types of data and quick processing. The primary consideration in algorithm creation is previous machine output.In this paper the prediction and detection of lung cancer is done by using many machine learning algorithms such as Naive Bayes, Random Forest, SVM, Linear Regression, K- nearest neighbors, KNN and Logistic Regression. By looking at the presentation of order computations, the major objective of this paper is to examining cellular breakdown in the lungs before it computations. We use convolution Neural Network Algorithms for this paper. And this survey will proceed by taking twitter data-sets which will predict on current rate of lung cancer. It is a type of Neural Networks. And his accuracy point is 92.45%.After the comparison of their results that are performed by using the different machine learning models and algorithms, that will results in the prediction of lung cancer.

Keywords- Convolution neural network, deep learning, Support vector machine, Lungs disease, data set, Decision tree, Random forest, Machine Learning.

I. INTRODUCTION

One of the technologies that the world finds to be most acceptable nowadays is machine learning. To create different types of algorithms that can be learned and observed from data, machine learning is used. The huge amount of data with correct estimation has a significant important role in the succeed of algorithms for prediction of lung cancer.The fundamental objective of this algorithms is its advanced accuracy and results in prediction of disease.It is a method to implement several health activities in the human body. Progressive clinical navigation refers to the cutting-edge application of AI to deal with clinical challenges. Machine learning's primary intent is to find trends in diverse health care activity data and carry out tasks in accordance with a data set. Machine learning produces more reliable and accurate results in health care systems. The patients of lung cancer undergo several symptoms such as chest pain, cough and shortness of breath.A pair of respiratory organs in the chest with a cone form called the lungs. As you breathe in, your lungs deliver oxygen to your body. As you exhale, they let off carbon dioxide, a waste product of the cells in your body.There are lobes in each lung. It has two lobes in the left lung. With three lobes, the right lung is slightly bigger. The lungs are encircled by a flimsy membrane known as the pleura. The trachea (windpipe) is connected to the right and left lungs by two tubes known as bronchi. Sometimes lung cancer also involves the bronchi. The interior of the lungs is made up of tiny tubes and air sacs called bronchi-oles. Small-cell lung cancer and non-small cell lung cancer are the two important category of lung cancer. And both of these is detected by appearance of cells in microscope. Non-small cell lung cancer is more dangerous than small cell lung cancer as it is classified as many variety and exceeds of lung cancer .Non-small cells cancer can be detected by chemotherapy radiation therapy and by combining both the therapy.Chemotherapy and radiation are more popularly used by patients in detection of lung cancer. In surgery they extract the cancer tissues and cells by surgical method. While in chemotherapy they take tablets injections and other types of drugs to reduce the effect of cancer. In Radiation therapy X rays are being used by doctor to kills the cells and tissues by using X-rays beams .This therapy does not allow the body to built and expand the insects inside the body .These drugs and injection can be inhaled by body through oral mouth or as pills.The probability of cancer stages is detected by performing many tests prior to thus therapy.90% of cancer is caused due to smoking, while others may be cancer heredity.We can predict and detect the existence and symptom of this disease at prior time by X-rays, CT scans etc.The four stages of cancer are based on the first stage, which is limited to lung cancer, the second stage, which is limited to chest cancer, the third stage, which features larger and more noticeable growths, and the fourth stage, during which the disease has spread to various body regions.

India and the remaining of the countries are experiencing an increase in lung cancer fatality, which are caused by the lung cancer disease worldwide. Lung cancer mortality is rising in India and the rest of the world, which is brought on by the disease globally.

II MOTIVATION

The main objective of choosing this domain is very essential by seeing all the current scenario and situation of the world in which mortality rate is rapidly every day. People are facing the problem of lung cancer and are not able to get proper guide and cure to rectify this diseases. So by observing and performing accurate operation and researches on lung cancer patients could get best result and boon by their health.The pain of lung cancer people

towards their inhaling condition and their struggle towards breath from fresh air is very sensitive that can attract everyone.As people who are suffering from lung cancer fells discomfort during pollution . Hence if researches are performed in positive manner then it can save many lives of many peoples and can decrease the fatality rate in the world. Hence this brings the main impact and motive for every researchers to save life on this earth by prediction of disease .

III LITERATURE SURVEY

[1] In this paper **Syed Saba Raouf** with his researchers worked on deep learning techniques and other neural network technologies.Their work is totally based on finding and predicting data rate as higher because production is lower for other work.[2]**Mohd. Atikur Rahman** processed studied on Naive Bayes Random Forest Logistic Regression and Multilayer Perceptron methodologies.After processing these methodologies lung prediction has increased the effective of cancer prediction and but by these results it increases the chances of errors .These models can be used in future reference also to achieve highest accuracy results.By this there is a decrease in inter –and infra-reader variations. [3] **Xi Wang** uses MIL methods Context –Aware Block Selection and VGG-16 in his research work. The main theme of this model is Histopathology image analysis that serves as the accurate and effective methodology in prediction of lung cancer and but to achieve this target all the image has to be verified is a difficult threat in the stage .This model is not suitable for different types of stages as it can detect one stage at a time.[4] **Amrita Vishwa** proposed in his study about the Support vector machine Random Forest, Hybrid model methodologies. SVM is suitable for data having more noise along with the large data sets. The working of the artificial neural network with numerical data makes it difficult to predict the lung cancer at early stage.[5] **Farhan Hassan khan** used Random forest MLP and Naïve Bayes in their research work. This model can predict the first stage of cancer that can be beneficial for doctors in the treatments and cure of disease. This model is totally based on CNN but its hard to achieve accurate result. As COX model is simple and effective to use and modify. These models can be used in future reference also to achieve highest accuracy results.[6]**Gang V Saji** worked on many models like CNN and GBM for investigating lung prior condition. The dose is taken by automate lung cancer by less radiation dose.Machine and deep learning has made certain dedication for the process of medical image analysis from correct prediction rate of lung cancer.[7]**Deepak Rawat** researched in his paper report about the Random Forest Naive Bayes and Decision trees. Machine learning algorithms helped in prediction of cancer at prior stage of disease .Their research work is totally based on category of data-sets. Random Forest achieved highest degree of success in accuracy with 97%. This model can be kept for future reference with large data sets when increased number of patients arrived at same insane.WSI can not be directly analyzed due to the immeasurable computational cost. [8] **Yusaku Nitta** in

his research work investigated on Neural network. Decision trees and In this study hyper-plane are used as effective tool .This hyper-plane is further categorized into four units like soft maximum kernel and hyper plane. The detection of disease by all these gave effective results in terms of accuracy .But this model is not applicable for future generation as it can not perform much accurate results with huge amount of data-sets.[9]**M Agrawal** and his scholar proceed the study of lung cancer prediction by taking use of random forest and decision trees. After performing their operation on different data-sets and dimensions they clarify that if lung is know at early-stage then it can be treatment and cured by doctors .Their research achieved highest accuracy result for Random Forest.But the disadvantage is that it is not applicable for large no of data-sets .This early detection can be beneficial and can reduce the mortality rate But the wrong detection of disease can harm the future generation.[10]**Lynch C. M.** and **De Carlo** studied on Random forest Linear Regression and Gradient boosted machine in their research work. The classification of data-sets is mostly preferred by 92% by taking use of ANN algorithms with all its featured architectures and all its ultimate yielded design.In their study they visualize on large detection of lung cancer with less life survival time with huge RSME.[11]**Radhika P R** and other researchers worked on many machine learning algorithms like decision tree and SVM for their study After that they worked on correct results of classifiers.All these are performed by high and increasing cells of lungs of the patients' body for their results.[12] **Dr A Nagarajan J Vasanth wason** with their research scholars worked on Random Forest and other machine learning algorithms .They used image analyzing of image and studied on statistical data-sets and matrices which helped doctors in the prediction of lung cancer at prior time by certain determined parameters .GLCM performed better classification of difficult data-sets then calculating techniques in the detection of lung cancer.This method gave best and accurate result for detection of disease. [13] **Senthil kumar** and **V Preethi** worked on Decision Trees Logistic Regression and SVM methodological. The classification and pattern recognition also use to embedded in the CAD system and it damages the lungs ability to supply oxygen to the blood.[14]**Chaskar Shashikant Rathore** investigated on Random forest Decision trees and K – Nearest neighbors It avoids time consumption and provide easy diagnosis.The performance of the Ada Boost classifier is effect with the different machine learning algorithms.[15]**Timor Kadir** and his researchers performed prediction of lung cancer by KNN. The easiest way to detect lung cancer is Neural network and genetic algorithms by the use of image. This tools is very effective to prediction the lung.These researchers worked on multiple filtering and small parts such as segments because they effect the body at early stage and may harms the patients body by other diseases.

IV. MACHINE LEARNING ALGORITHM FOR DETECTING LUNG CANCER

Linear Regression - It is a supervised machine learning language in which retrogression is performed.All the independent values and parameters are observed by activator in this model .This shows connection between parameters. Their behaviour towards their model depends the type of model used by them . And the model is used according to their coordination between themselves and their the amount of parameters used by them. Hence its shows a straight connection between them And that is why it is called Rectilinear retrogression. As the figure shows that X is input of work experience and Y is affair which denotes payment of the data.The stylish fit line is called the retrogression wind in the figure.
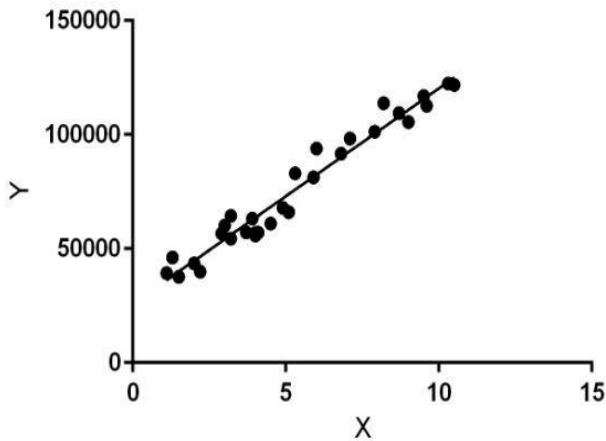
.

Fig.1: Linear Regression

Random Forest - This is a supervised learning language who is based on ensemble rate which is a combined version of all different algorithms or its original form of significant model.It is a combination of many random forest algorithms that is called forest and hence later it is trained by bagging . These are mostly used for solving all type of problems like regression and parenthesis. If it has many trees then the accuracy and efficiency is high .The training set of these trees is observed by the result of decision trees. Decision tree is less accurate then random forest because it uses all types of trees algorithms .
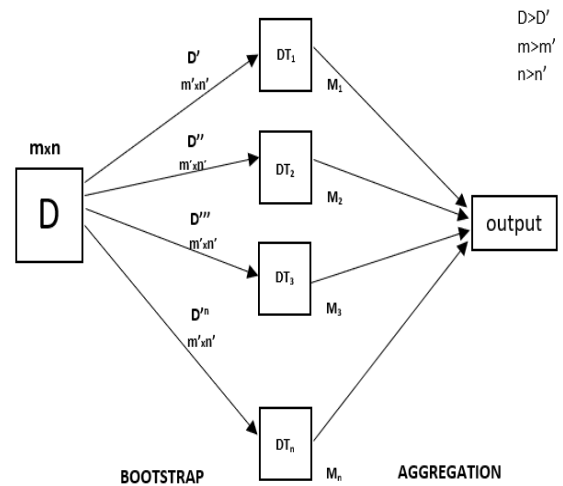

Fig.2: Random Forest

K Neighbors- K-Nearest Neighbors is very effective and simpler algorithms as compared to other algorithms therefore it is called base of other algorithms. It is essential supervised learning algorithm that can be beneficial used for data process find pattern of the data .Its working rule is also very simple which is totally based on data point that are designated by neighbors. Its also makes clusters by Euclidean formulation .This is very non-variable so that it can be beneficially applied on current situation of disease as it is not depend on frequency of distributed data-sets. The merit of this algorithms is very high as compared to other computational time resources .
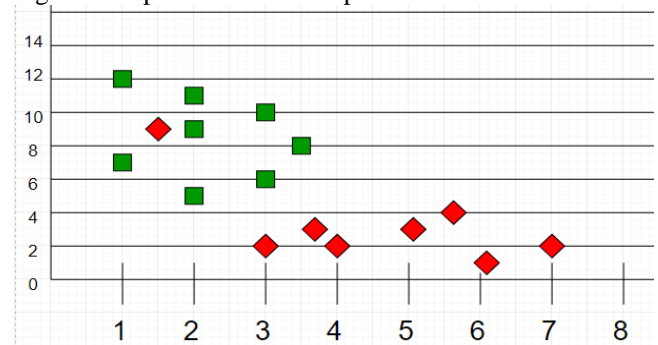

Fig.3: K-Nearest Neighbors

Support Vector Machine (SVM) -

Support Vector Machine is based on both types of problems such as classification and regression .As we are building our model in prediction of lung cancer then it is very effective and beneficial for our research .Data-sets that are being used in this research is non linear type. Hence it is supportive for detecting accuracy results for lung prediction.
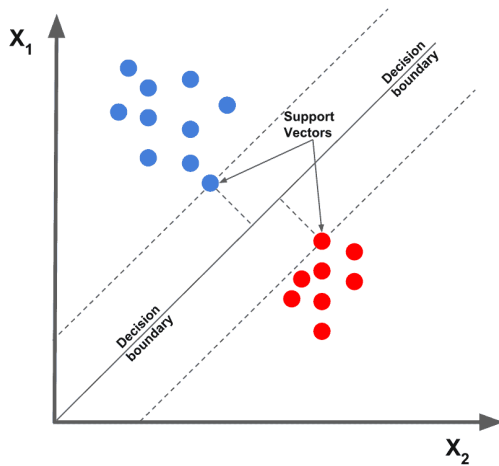
Fig 4: SVM

V. CONCLUSION

The proposed framework is appropriate to more both harmless and horrible excrescences all the more properly. In this way, it tends to be reasoned that the SVM utilized for excrescence section have delicacy of 100 that can manage difficult gatherings. Expectation and its initial discovery of the cellular breakdown in the lungs can be broadened future by utilizing Support Vector Machine and some factual strategy to anticipate its beginning phases of disease and furthermore precisely. In this study, different variety that are already utilizing AI calculations and algorithms are used to forecast cellular breakdown in the lungs.

REFERENCES

[1] Raoof, Syed Saba, M. A. Jabbar, and Syed Aley Fathima. "Lung Cancer prediction using machine learning: A comprehensive approach." In *2020 2nd International conference on innovative mechanisms for industry applications (ICIMIA)*, pp. 108-115. IEEE, 2020.

[2] Bharati, Subrato, Mohammad Atikur Rahman, and Prajoy Podder. "Breast cancer prediction applying different classification algorithm with comparative analysis using WEKA." In *2018 4th International Conference on Electrical Engineering and Information & Communication Technology (iCEEiCT)*, pp. 581-584. IEEE, 2018.

[3] Wang, Xi, Hao Chen, Caixia Gan, Huangjing Lin, Qi Dou, Efstratios Tsougenis, Qitao Huang, Muyan Cai, and Pheng-Ann Heng. "Weakly supervised deep learning for whole slide lung cancer image analysis." *IEEE transactions on cybernetics* 50, no. 9 (2019): 3950-3962.

[4] Thallam, Chinmayi, Aarsha Peruboyina, Sagi Sai Tejasvi Raju, and Nalini Sampath. "Early stage lung cancer prediction using various machine learning techniques." In *2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, pp. 1285-1292. IEEE, 2020.

[5] Faisal, Muhammad Imran, Saba Bashir, Zain Sikandar Khan, and Farhan Hassan Khan. "An evaluation of machine learning classifiers and ensembles for early stage prediction of lung cancer." In *2018 3rd international conference on emerging trends in engineering, sciences and technology (ICEEST)*, pp. 1-4. IEEE, 2018.

[6] Saji, Ganga V., Thasneem Vazim, and Sumod Sundar. "Deep Learning Methods for Lung Cancer Detection, Classification and Prediction-A Review." In *2021 Fourth International Conference on Microelectronics, Signals & Systems (ICMSS)*, pp. 1-5. IEEE, 2021.

[7] Rawat, Deepak. "Validating and Strengthen the Prediction Performance Using Machine Learning Models and Operational Research for Lung Cancer." In *2022 IEEE International Conference on Data Science and Information System (ICDSIS)*, pp. 1-5. IEEE, 2022.

[8] Khalsan, Mahmood, Lee R. Machado, Eman Salih Al-Shamery, Suraj Ajit, Karen Anthony, Mu Mu, and Michael Opoku Agyeman. "A survey of machine learning approaches applied to gene expression analysis for cancer prediction." *IEEE Access* 10 (2022): 27522-27534.

[9] Dang, Anh. "Cancer Prediction using Machine Learning Algorithms."

[10] Lung Cancer Survival Prediction Using Supervised Machine-learning Classification Techniques, International Journal of Medical Informatics, JL. Patel and R. Sharpe 108, 1-8.

[11] Radhika, P. R., Rakhi AS Nair, and G. Veena. "A comparative study of lung cancer detection using machine learning algorithms." In *2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT)*, pp. 1-4. IEEE, 2019.

[12] Nagarajan, Dr A., and J. Vasanth Wason. "Machine learning approach to predict lung cancer using CT scan images." *International Journal of Advanced Trends in Computer Science and Engineering* 8, no. 6 (2019): 2974-2976.

[13] Kavethanjali, V., S. Preethi, and V. Vasanthapriya. "Lung–Pleura Carcinoma Detection Using Machine Learning." In *2021 3rd International Conference on Signal Processing and Communication (ICPSC)*, pp. 294-298. IEEE, 2021.

[14] Ingle, Kshitija, Uttam Chaskar, and Shashikant Rathod. "Lung Cancer Types Prediction Using Machine Learning Approach." In *2021 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT)*, pp. 01-06. IEEE, 2021.

[15] Kadir, Timor, and Fergus Gleeson. "Lung cancer prediction using machine learning and advanced imaging techniques." *Translational lung cancer research* 7, no. 3 (2018): 304