IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 9, Issue 6, June 2021

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.542

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 9, Issue 6, June 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0906221 |

Lung Cancer Detection Using Machine Learning

Priyanka Agawane, Manisha Khebale, Yogendra Patil

UG Student, Dept. of Computer, CCIS, JSPM's, Bhivarabai Sawant Institute of Technology & Research, Wagholi, Pune, India

UG Student, Dept. of Computer, CCIS, JSPM's, Bhivarabai Sawant Institute of Technology & Research, Wagholi, Pune, India

Assistant Professor, Dept. of Computer, CCIS, JSPM's, Bhivarabai Sawant Institute of Technology & Research, Wagholi, Pune, India

ABSTRACT: Lung cancer is one of the main cause of the death and health issue in many countries with a 5- year survival rate of only 10–16%. In this project we use machine learning algorithms to diagnose a cancer and start treatment in early stages. We use KNN & Decision Tree algorithms to predict the accuracy of the cancer. In this project we use scikit-learn libraries like sklearn and pandas to predict and classify the dataset of the lung cancer patients. Slicing the dataset and feature scaling options are used to train the dataset. After that we use confusion matrix , f1 score and accuracy score to predict the accuracy of the result.

KEYWORDS: Classifier, Lung Cancer, Machine Learning

I. INTRODUCTION

Lung cancer is one of the main cause of the death and health issue in many countries with a 5- year survival rate of only 10–16%. People who smoke have the greatest risk of lung cancer, though lung cancer can also occur in people who have never smoked. The risk of lung cancer increases with the length of time and number of cigarettes you've smoked. If you quit smoking, even after smoking for many years, you can significantly reduce your chances of developing lung cancer. In this project we use machine learning algorithms to diagnose a cancer and start treatment in early stages. With the expected increase in the number of preventive/early detection measures, scientists are working in computerized solutions that help alleviate the work of doctors, improve diagnostics' precision by reducing the subjectivity factor, speedup the analysis and reduce medical costs. In order to detect malignant nodules, specific features need to be recognized and measured. Based on the detected features and their combination, cancer probability can be assessed. However, this task is very difficult, even for an experienced medical doctor, since nodule presence and positive cancer diagnosis are not easily related.

In this project we have use knn & decision tree algorithms to predict the accuracy of the cancer. In this project we use scikit-learn libraries like sklearn and pandas to predict and classify the dataset of the lung cancer patients. Slicing the dataset and feature scaling options are used to train the dataset. After that we use confusion matrix , f1 score and accuracy score to predict the accuracy of the result.

II. RELATED WORK

In 2018, 1.76 million people worldwide died of lung cancer. Most of these deaths are due to late diagnosis, and early-stage diagnosis significantly increases the likelihood of a successful treatment for lung cancer. Machine learning is a branch of artificial intelligence that allows computers to quickly identify patterns within complex and large datasets by learning from existing data. Machine-learning techniques have been improving rapidly and are increasingly used by medical professionals for the successful classification and diagnosis of early-stage disease. They are widely used in cancer diagnosis. In particular, machine learning has been used in the diagnosis of lung cancer due to the benefits it offers doctors and patients. In this context, we performed a study on machine-learning techniques to increase the classification accuracy of lung cancer with 32 × 56 sized numerical data from the machine learning repository web site

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 9, Issue 6, June 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0906221 |

of the university of california, irvine. In this study, the precision of the classification model was increased by the effective employment of pre-processing methods instead of direct use of classification algorithms. Nine datasets were derived with pre-processing methods and six machine-learning classification methods were used to achieve this improvement. The study results suggest that the accuracy of the k-nearest neighbors algorithm is superior to random forest, naïve bayes, logistic regression, decision tree, and support vector machines. The performance of pre-processing methods was assessed on the lung cancer dataset. The most successful preprocessing methods were z-score (83% accuracy) for normalization methods, principal component analysis (87% accuracy) for dimensionality reduction methods, and information gain (71% accuracy) for feature selection methods

III. PROPOSED ALGORITHM

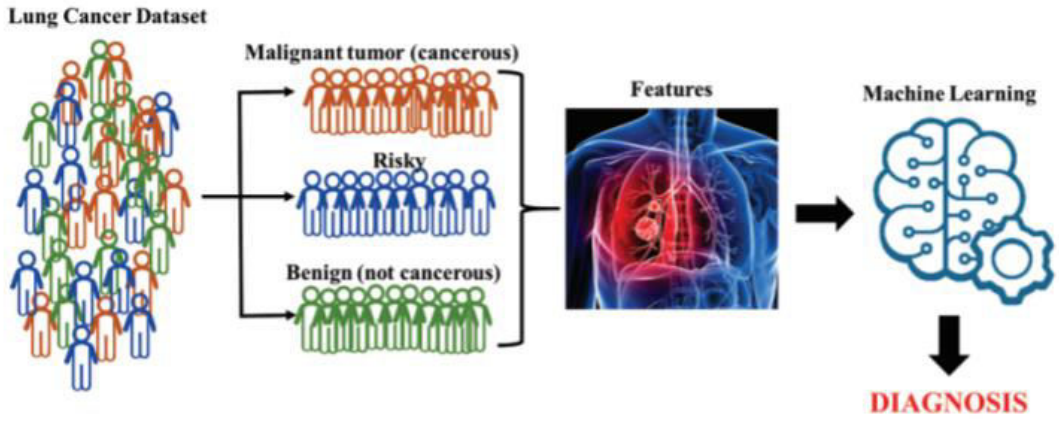

Fig 1 Architectural Diagram Of Proposed Model

IV. PSEUDO CODE

Feature selection is applied to reduce the number of features in many applications where the data have hundreds or thousands of properties. The main idea is to find globally the least reduction or, in other words, the smallest set of features that represent the most important characteristics of the original set of features [16]. The choice of methods for the machine-learning prediction system is important because there are many machine-learning algorithms used in practice for particular purposes. For instance, random forest (RF) works with the logic of increasing the accuracy of results by deriving multiple decision trees while k-nearest neighbors (k-NN) uses similarities to find neighbors when classifying by majority vote. Naïve Bayes (NB) maintains the most appropriate classification by preserving the dependence of the qualifications on a particular class. Logistic regression (LR) finds the dependent and independent relationships between the variables affected by the dependent variables. Decision tree (DT) is the preferred learning method with a created tree structure since it is faster, easier to interpret, and is more effective. Support vector machines (SVMs) work with hyperplanes to separate data classification into a multidimensional space [4,17]. k-NN has been reported to give the best results for machine-learning algorithms applied to the histopathological classification of non-small cell carcinomas. The DT algorithm was reported to give the least favorable results [18]. This study aims to develop predictive models to diagnose lung cancer disease based on a customized machine-learning framework. This approach involves examining the different degrees of success of these models and analyzes their generally valid classification performances according to measurement metrics. In this context, this study consists of three modules. The first module is based on the application of the data pre-processing techniques (dimensionality reduction methods, normalization techniques, and feature selection methods) to the lung cancer dataset (LCDS), which is taken from the Machine Learning Repository website of the University of California, Irvine (UCI). The second module focuses on the demonstration and discussion of the performance of the machine-learning algorithms (RF, k-NN, NB, LR, DT, and SVMs). The third module looks at the results of all the performance measurement metrics and performs validation analysis. The aforementioned methods are widely used in diagnosis and analysis to make decisions in different areas of medicine and research.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 9, Issue 6, June 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0906221 |

V. CONCLUSION AND FUTURE WORK

The lung cancer detection system using the machine learning technique is much efficient and gives the betterment result to the radiologist and assist them. This enhances with the additional features for upgrading in the future. On this processing system to support the radiologist to detect the affected patients as accurate as the result. Machine learning is the key to enabling Artificial Intelligence and the future of healthcare is data-driven. Big data and machine learning have a tremendous potential in the healthcare field. All these technologies are not only improving treatment and diagnosis options, they also have the potential to take control of their own health by empowering individuals. With the help of advanced analytics, artificial intelligence and machine learning some of the most exciting advances are coming about in healthcare. Advances in AI interfaces, personalized medicine, predictive healthcare and advances in diagnostics all come down to the application of machine learning to help patients have access to smarter healthcare.

REFERENCES

1. https://archive.ics.uci.edu/ml/dataset/Lung+cancer. Accessed 12 Feb 2020
2. WHO Deaths by cause, sex and mortality stratum, World Health Organization. https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death. Accessed 25 Jan 2020
3. Ada, R.K.: Early detection and prediction of lung cancer survival using neural network classifier (2013)
4. Alcantud, J.C.R., Varela, G., Santos-Buitrago, B., Santos-Garcia, G., Jimenez, M.F.: Analysis of survival for lung cancer resections cases with fuzzy and soft set theory in surgical decision making. PLoS ONE **14**(6), e0218283 (2019)
5. Asuntha, A., Srinivasan, A.: Deep learning for lung cancer detection and classification. Multimedia Tools Appl. **79**, 1–32 (2020)
6. Bhatia, S., Sinha, Y., Goel, L.: Lung cancer detection: a deep learning approach. In: Bansal, J.C., Das, K.N., Nagar, A., Deep, K., Ojha, A.K. (eds.) Soft Computing for Problem Solving. AISC, vol. 817, pp. 699–705. Springer, Singapore (2019).
7. Bhuvaneswari, P., Therese, A.B.: Detection of cancer in lung with k- nn classification using genetic algorithm. Procedia Mater. Sci. **10**, 433–440 (2015)
8. Chaubey, N.K., Jayanthi, P.: Disease diagnosis and treatment using deep learning algorithms for the healthcare system. In: Applications of Deep Learning and Big IoT on Personalized Healthcare Services, pp. 99–114. IGI Global (2020)
9. Ganggayah, M.D., Taib, N.A., Har, Y.C., Lio, P., Dhillon, S.K.: Predicting factors for survival of breast cancer patients using machine learning techniques. BMC Med. Inform. Decision Making **19**(1), 48 (2019)
10. Hachesu, P.R., Moftian, N., Dehghani, M., Soltani, T.S.: Analyzing a lung cancer patient dataset with the focus on predicting survival rate one year after thoracic surgery. Asian Pacific J. Cancer Prevention: APJCP **18**(6), 1531 (2017)
11. Hosseinzadeh, F., KayvanJoo, A.H., Ebrahimi, M., Goliaei, B.: Prediction of lung tumor types based on protein attributes by machine learning algorithms. SpringerPlus **2**(1), 238 (2013)
12. Hussein, S., Kandel, P., Bolan, C.W., Wallace, M.B., Bagci, U.: Lung and pancreatic tumor characterization in the deep learning era: novel supervised and unsupervised learning approaches. IEEE Trans. Med. Imag. **38**(8), 1777–1787 (2019)
13. Jacob, D.S., Viswan, R., Manju, V., PadmaSuresh, L., Raj, S.: A survey on breast cancer prediction using data mining techniques. In: 2018 Conference on Emerging Devices and Smart Systems (ICEDSS), pp. 256–258. IEEE (2018)
14. Jakimovski, G., Davcev, D.: Using double convolution neural network for lung cancer stage detection. Appl. Sci. **9**(3), 427 (2019)
15. Kadir, T., Gleeson, F.: Lung cancer prediction using machine learning and advanced imaging techniques. Transl. Lung Cancer Res. **7**(3), 304 (2018)
16. Kohad, R., Ahire, V.: Application of machine learning techniques for the diagnosis of lung cancer with ant colony optimization. Int. J. Comput. Appl. **113**(18), 34–41 (2015)
17. Kourou, K., Exarchos, T.P., Exarchos, K.P., Karamouzis, M.V., Fotiadis, D.I.: Machine learning applications in cancer prognosis and prediction. Comput. Struc. Biotechnol. J. **13**, 8–17 (2015)
18. Krishnaiah, V., Narsimha, G., Chandra, D.N.S.: Diagnosis of lung cancer prediction system using data mining classification techniques. Int. J. Comput. Sci. Inf. Technol. **4**(1), 39–45 (2013)

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 9, Issue 6, June 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0906221 |

19. Li, X., Hu, B., Li, H., You, B.: Application of artificial intelligence in the diagnosis of multiple primary lung cancer. Thoracic Cancer **10**(11), 2168–2174 (2019)
20. Lynch, C.M., et al.: Prediction of lung cancer patient survival via supervised machine learning classification techniques. Int. J. Med. Inform. **108**, 1–8 (2017)
21. Monkam, P., Qi, S., Ma, H., Gao, W., Yao, Y., Qian, W.: Detection and classification of pulmonary nodules using convolutional neural networks: a survey. IEEE Access **7**, 78075–78091 (2019)
22. Murty, N.R., Babu, M.P.: A critical study of classification algorithms for lungcancer disease detection and diagnosis. Int. J. Comput. Intell. Res. **13**(5), 1041–1048 (2017)
23. Paing, M.P., Hamamoto, K., Tungjitkusolmun, S., Pintavirooj, C.: Automatic detection and staging of lung tumors using locational features and double-staged classifications. Appl. Sci. **9**(11), 2329 (2019)
24. Patel, D., Shah, Y., Thakkar, N., Shah, K., Shah, M.: Implementation of artificial intelligence techniques for cancer detection. Augmented Human Res. **5**(1), 6 (2020)
25. Pradeep, K., Naveen, N.: Lung cancer survivability prediction based on performance using classification techniques of support vector machines, c4. 5 and naive bayes algorithms for healthcare analytics. Procedia computer science **132**, 412–420 (2018)
26. Shakeel, P.M., Tolba, A., Al-Makhadmeh, Z., Jaber, M.M.: Automatic detection of lung cancer from biomedical data set using discrete adaboost optimized ensemble learning generalized neural networks. Neural Comput. Appl. **32**(3), 777–790 (2020)

INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 7.542

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details