International Journal of Data Informatics and Intelligent Computing (IJDIIC)
Vol. 2, No. 3, September 2023, pp. 47~56
DOI: 10.59461/ijdiic.v2i3.73

Machine Learning Techniques for Lung Cancer Risk Prediction Using Text Dataset

Kumar Mohan[1], Bhraguram Thayyil[1]
[1]Department of Information Technology, University of Technology and Applied Sciences-Shinas, Al Aqar, Oman

Article Info

Article history:

Received July 12, 2023
Revised September 17, 2023
Accepted September 22, 2023

Keywords:

Machine learning
Prediction
Lung Cancer
Random forest

ABSTRACT

The early symptoms of lung cancer, a serious threat to human health, are comparable to those of the common cold and bronchitis. Clinical professionals can use machine learning techniques to customize screening and prevention strategies to the unique needs of each patient, potentially saving lives and enhancing patient care. Researchers must identify linked clinical and demographic variables from patient records and further preprocess and prepare the dataset for training a machine-learning model in order to properly predict the development of lung cancer. The goal of the study is to develop a precise and understandable machine learning (ML) model for early lung cancer prediction utilizing demographic and clinical variables, as well as to contribute to the growing field of medical research ML application that may improve healthcare outcomes. In order to create the most effective and precise predictive model, machine learning techniques like Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, K-Nearest Neighbor (KNN), and Naive Bayes were utilized in this article.

Corresponding Author:

Kumar Mohan
Department of Information Technology
University of Technology and Applied Sciences-Shinas
Al Aqar
Oman
Email: kumar.mohan@shct.edu.om

1. INTRODUCTION

In today's world, lung cancer causes more deaths from cancer-related causes than any other cancer, like breast, prostate, and colon cancers. Early lung cancer diagnosis and treatments have satisfiable consideration and a positive impact on patient outcomes and mortality rates. The disease was frequently discovered at its advanced stages in many lung cancer cases since existing screening techniques are not always reliable. For the early diagnosis of lung cancer, machine-learning techniques present a viable approach [1]. Machine learning models may reliably forecast a patient's likelihood of acquiring lung cancer by identifying patterns and risk factors linked to the disease by evaluating vast databases of patient records. Using medical imaging data from CT scans and X-rays, machine-learning models have been created in recent years to help with the early identification of lung cancer. However, these techniques can be expensive and time-consuming. In this article, we explore the application of machine learning techniques to estimate the risk of lung cancer using text datasets. The model specifically looks at electronic health records (EHRs) that include patient data such as demographics, clinical notes, and medical history. We need to properly forecast a patient's risk of acquiring lung cancer by extracting pertinent elements from these text datasets and building a machine-learning algorithm. This section tries to demonstrate an overall review of studies that have analyzed genetic expression data in cancer. This includes many matrix methods like (Confusion Matrix, Accuracy, Precision, Recall, and F1 Score), and Algorithm methods such as (Logistic Regression, Decision Tree, Random Forest, Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Naive Bayes) to discover genes from the samples and using these expression signatures to develop a cancer prediction model.

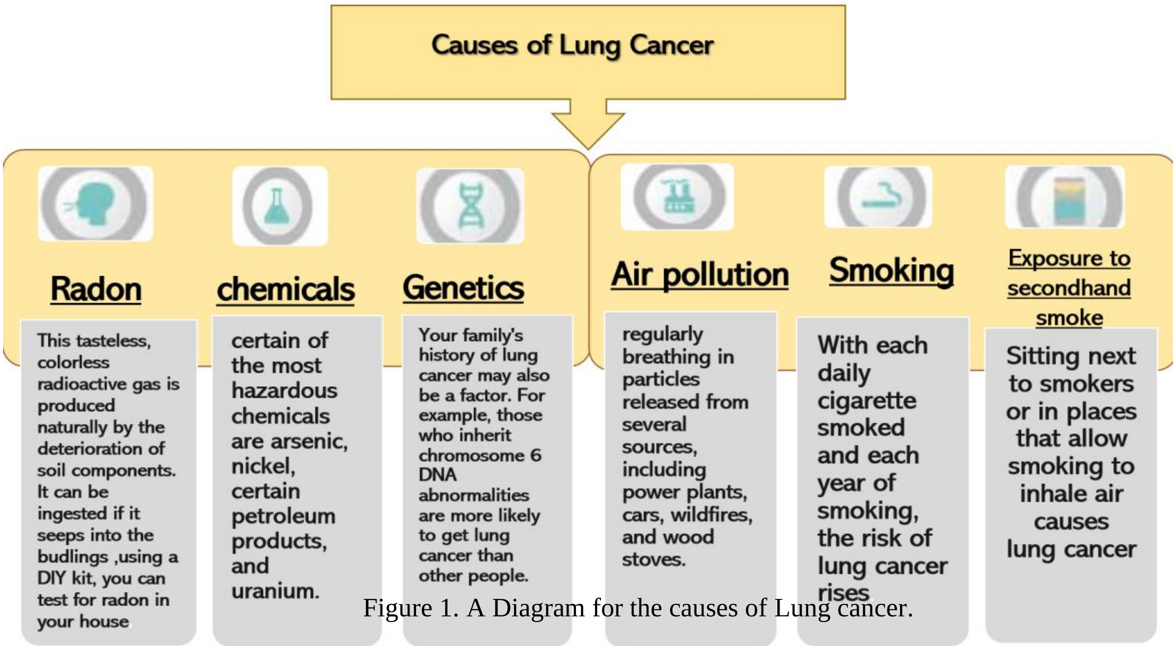


Figure 1. A Diagram for the causes of Lung cancer.

Figure 1 shows the causes of lung cancer, frequently brought into the human body by various lifestyle habitats like smoking. Both smokers and people who have been exposed to secondhand smoke also. A few wounds might happen in non-smokers, for example, individuals presented to radon gas, openness to cancer-causing agents, and a family background of cellular breakdown in the lungs.

2. LITERATURE REVIEW

Firstly, the article [4] presented features based on deep learning and performance-wise, the sample beat the conventional machine learning strategy. For the deep learning sample, the method obtained a value of 0.5 for the regression strategy and a maximum accuracy of 71.18 per cent for the classification approach. However, the regression sample RMSE for the conventional machine learning models remained constant at 14.87 and 61.12 per cent, respectively. In the article [5], about 34% of lung malignancies are seen in pulmonary nodules, which are rounded or unequal lesions in the lungs. As a result, pulmonary nodule diagnosis is crucial for the early discovery of lung cancer. The findings of the experiment demonstrate that the convolutional neural network-based method for identifying and detecting lung cancer has a greater level of accuracy. In addition, the article [6] considers Web of Science (WoS) articles and book review databases, which leads the researchers to use bibliometrics to conduct a quantitative analysis of surveys produced in the 24 nations that are at the top of the world for cancer research. The analysis includes 32,161 research papers on lung cancer from 2085 distinct journals. Around 5.6% of all cancer research was focused on lung cancer in 2013, and around 1.2% increase since 2004. The information shows that, despite the significant clinical, societal, and financial consequences linked with lung cancer, the amount of global research output trails substantially back of malignancies.

In article [7], the prognosis for locally progressing, incurable non-small cell lung cancer steadily produces better results by the inclusion of chemotherapy to radical radiotherapy and new radiotherapy treatments. With current drugs being better tolerated and resulting in an improvement in quality of life, chemotherapy gives moderate survival improvements for individuals with non-small cell lung cancer. Article [8] presented better results of the median (interquartile range) age of the 17 322 NSCLC patients, where it included 68 (61-74) and 13 361 (77.1%) of them were white in the evaluation criteria. Ten thousand two hundred seventy-three stage I tumours (59.3% of all cancers) and 11 985 adenocarcinomas (69.2% of all tumours) accounted for the bulk of tumours. A follow-up period of 24 (10–43) months was the median (interquartile range). During the follow-up period, 3119 patients passed away because of lung cancer.

The article [9] presented that Lung cancer is a serious cause of cancer-related fatalities in the whole world. The lung cancer mortality rate is reduced by 20% to 24% when lung cancer is screened for on a chest computed tomography (CT) scan. In article [10], With the use of CT scan pictures, a detailed scanning method excavated using a new technical strategy, the challenge of identifying benign and malignant kinds of lung cancer may be solved, aiding in the medical community's ability to diagnose lung cancer more easily. According to the system trial, the accuracy rate for predicting whether a lung cancer diagnosis is benign or malignant is 83.33%. The article [11] concentrates more on age, CT and PET risk, and the risk of OS has been used to predict while the PET risk has a negative impact on OS with an HR of 0.67 (covering impact). The CT danger has a positive impact on OS with an HR of 1.35 (an enhanced risk), and the article tries to prove that age has a negligible impact.

In the article [12], the authors consider the population level of the US and the NSCLC mortality rate, which fell considerably between 2013 and 2016, whereas survival after diagnosis was shown to increase significantly. This study predicts a decrease in incidence associated with medical advancements, notably the usage and approvals of targeted medicines. Artcle [13] projects a technical attempt to create an automated system for detecting lung cancer. The methods used in this study produced a hospital database accuracy of 92%. This technology attempts to improve the precision and speed of lung cancer detection systems. Additionally, it aids in earlier cancer detection.

In the article [14], the patients with early-phase inoperable lung cancer now receive stereotactic ablative radiotherapy (SABR) as a reasonable level of care. The majority of these patients are seniors and have mutations in the EGFR (epidermal growth factor receptor). To determine how ageing and the EGFR switch affected therapy outcomes and toxicity, the article looked at the medical records of 71 patients who got SABR at Taipei Veterans General Hospital between 2015 and 2021 and had incurable, early-stage non-small cell lung cancer (NSCLC). Stereotactic body radiotherapy (SBRT) or stereotactic ablative radiotherapy (SABR) can finish the treatment in one to two weeks. These findings show that SABR efficacy and safety in patients with early-stage, incurable NSCLC were unaffected by age or EGFR mutation status. 37 (52.1%) of these patients were 80 years of age or older, 50 (70.4%) had T1 illnesses, and 21 (29.6%) had T2 diseases. 33 (46.5%) patients' EGFR mutation status was known, and 16 (51.5%) of those patients had a mutation.

In the article [15], the authors presented numerous advantages of using early cancer diagnosis to predict survival rates. Using deep learning methods, a strategic approach has been developed, and a number of survival prediction models for lung cancer patients have been developed to tackle the classification and regression of cancer survival issues presented. We have analyzed the presentation across three of the most famous profound learning structures - Counterfeit Brain Organizations (ANN), Convolutional Brain Organizations (CNN), and Intermittent Brain Organizations (RNN) while looking at the performance of profound learning models as opposed to the conventional AI sample. In both classification and regression methods, the deep learning models outperformed the traditional machine learning models in terms of performance. The researchers obtained a 0.5 rate for the regression approach and a maximum accuracy of 71.18 per cent for the classification method for the deep learning models. The RMSE in the regression model for the traditional machine learning models remains unchanged at 14.87 per cent and 61.12 per cent, respectively.

In article [16], the authors falicitates an early detection mechanism by the study of recent years. AI and machine learning have been studied and utilized in the early detection of this disease. New methods have been developed by combining data detection and processing of biomedical images. On a lung cancer disease dataset, machine learning algorithms were used to classify images and calculate accuracy, sensitivity, and other metrics in this paper. The initial stage of lung cancer has been analyzed using the K-NN, Random Forest, and SVM algorithms. It may take the resultant decision of lung cancer affected rate that it has a higher death rate than other types of cancer. It is possible to treat the disease promptly and save lives if it is detected earlier. CT imaging has been adopted as a wide-margin, more trusted, and exact approach to distinguishing disease improvement in lungs since uncovering any thought unsuspected cellular breakdown in the lung nodules is possible as well. However, since the intensity of CT images varies, precise detection remains a significant challenge. Computer-aided diagnosis (CAD) has emerged as a major supporting strategy in the fight against this issue. In addition, research is being done in this area of lung cancer detection in order to achieve 100% detection accuracy. In this article, cutting-edge innovations utilized for the expectation of cellular breakdown in the lungs utilizing pictures from CT filters are talked about, and an examination has led them to distinguish the top method. Three classifiers (K-NN, Irregular woods and SVM) were used as the resultant variations and based on the outcomes and comparison, the authors determined that Random Forest, followed by SVM, with an accuracy of 82.1 per cent, produced top results among the three algorithms.

In article [17], a survey has been done on the study to give an outline of AI-based approaches that reinforce the changing parts of cellular breakdown in the lung's conclusion and treatment, including early location, assistant finding, forecast expectation and immunotherapy practice. Around 2.20 million new patients are diagnosed with a cellular breakdown in the lungs every year, and 75% of them pass on in the span of five years of analysis. The intricacy of cancer cells that cause medication resistance and the high intra-tumor heterogeneity (ITH) make treating cancer more challenging. The continual advancement of cancer research technology over the past few decades has led to the formation of multiple clinical, medical imaging, and genomic databases as a result of numerous big collaborative cancer projects. These databases make it simpler for researchers to look into full patterns of lung cancer, from diagnosis to treatment and responses to clinical consequences. Current research on -omics analysis, encompassing genomes, transcriptomics, proteomics, and metabolomics, has particularly improved our research tools and capacities. Chest screening with low-dose calculated tomography (CT) is the main technique for monitoring people who are at risk for lung cancer. The goal of a computer-aided diagnosis (CAD) system is to increase the effectiveness of diagnostics by assisting physicians with the interpretation of medical imaging data.

Article [18] presented a prototype model for the treatment of lung cancer, which can be created without endangering the environment by utilizing the most recent developments in computational intelligence. The system saves time and money since it may cut down on the number of resources that are squandered and the amount of labour required to do manual tasks. The detection process from the lung cancer dataset was optimized using a machine learning model built on support vector machine (SVM) architecture. An SVM classifier is utilized to categorize lung cancer patients according to their symptoms, and Python is used to enhance the model's implementation. We evaluated the effectiveness of our SVM model using a variety of criteria to build model based architecture. Pneumonic illness accounted for over 13% of all malignant growth analyses in the US in 2015, according to several studies. According to the American Cancer Society, 27% of all cancer-related fatalities are thought to be caused by lung cancer. As a result, it's important to properly assess and monitor lung nodules during their development stages. The aim of this survey is to explore the growth and progression of cancer using the ML and DL methodologies for growth and progression prediction.

In conclusion, the literature review provides a comprehensive overview of lung cancer, its causes, diagnosis, and treatment options. It highlights the importance of prevention and early detection in upgrading outcomes for lung cancer patients. Moreover, researchers anticipated identifying new therapy options and upgrading the overall survival rates for lung cancer patients.

3. METHODOLOGY

A combination of data collection, preprocessing, model development, validation, optimization, and interpretation have been used in this article as a research method for machine learning methods for the prediction of lung cancer risk using text datasets in order to create an accurate and dependable predictive model for this crucial health issue. This research used a mixed approach (quantitative and qualitative) and used secondary data for better results. This experimental research ran through various health-based criteria and formulated an approach with a mixed algorithmic model.

Proposed System: After preprocessing, the data will be ready to use in model creation. The build model procedure requires the use of machine learning algorithms and a preprocessed dataset. Some of the techniques employed include LR, DT, and RF classification. There are a certain number of factors that have been evaluated and tested in this article. The diagrammatic depiction of the designed system is displayed in Figure 2. The following subsections cover all of the block diagram's elements in detail [20]. The technique of text data gathering that might be adopted is survey research.

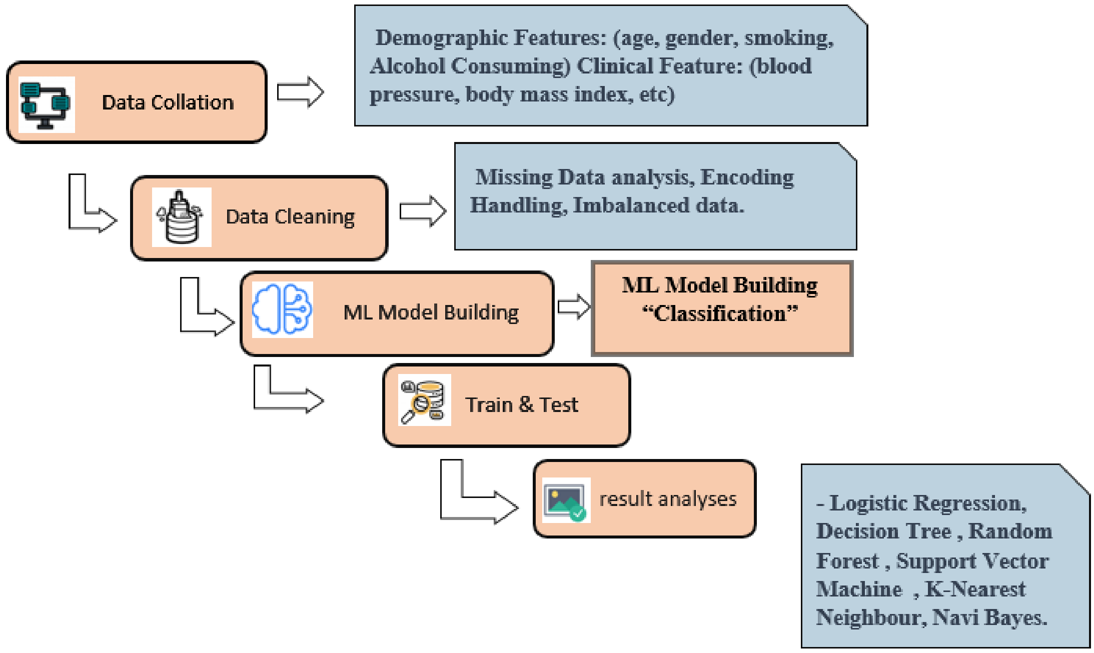


Figure 2. A Diagram for production of Lung cancer.

There are several reasons that should be considered while survey research might be used in lung cancer research:

Recognizing risk factors: Survey research may be used to collect information on habits, way of life, and environmental elements that may raise the risk of lung cancer. This information might be used by researchers to spot trends and disease-related risk factors.

Understanding patient experiences: Survey research can also be utilized to comprehend patient experiences after receiving a lung cancer diagnosis. Researchers can learn about symptoms, treatment experiences, and quality of life concerns by interviewing patients.

Evaluation of interventions: Survey research can be used to assess the efficacy of lung cancer prevention or treatment programs. For instance, surveys might be used by researchers to collect information on lung cancer screening programs or initiatives to help smokers quit.

In general, survey research is highly useful method for obtaining information on many aspects of lung cancer, which can serve to guide preventative and treatment plans [21].

3.1. Data preprocessing

Because of missing values and/or noisy data, the quality of the raw data might be worse than the quality of the final forecast. In order to make data more suitable for mining and analysis, preprocessing is required, and a thorough reduction of redundant values, characteristic selection, and data discretization is performed. Class balancing using a resampling approach is another component of data preparation. We used the structural specification of SMOTE in the suggested framework to alleviate the uneven division of the participants across the lung cancer and non-lung cancer groups. More particularly, the participants were evenly distributed because the minority group, in the lung cancer condition, was excessive in certain conditions. Additionally, no data inforrcement or dropping was used because there were no missing or null values.

Each participant's gender and the age group to which they belong are shown in Figure 3, together with the distribution of students in each class. With regard to the lung cancer class specification, the second figure part shows that a sizable portion of the participants are over 80 years old, while the second, most common age range is between 30-82. Additionally, we can see from this figure that lung cancer primarily affects elderly people. In contrast, In the left figure, a sizable portion of the participants are over 75 years old, while the second most common age range is between 38 and 79. Additionally, we can see from this figure that lung cancer primarily affects elderly people. According to Figure 3, there were around 27% and 36% more women than men affected by lung cancer. That indicates although lung cancer still affects both men and women, men are 9% more likely to develop the disease.

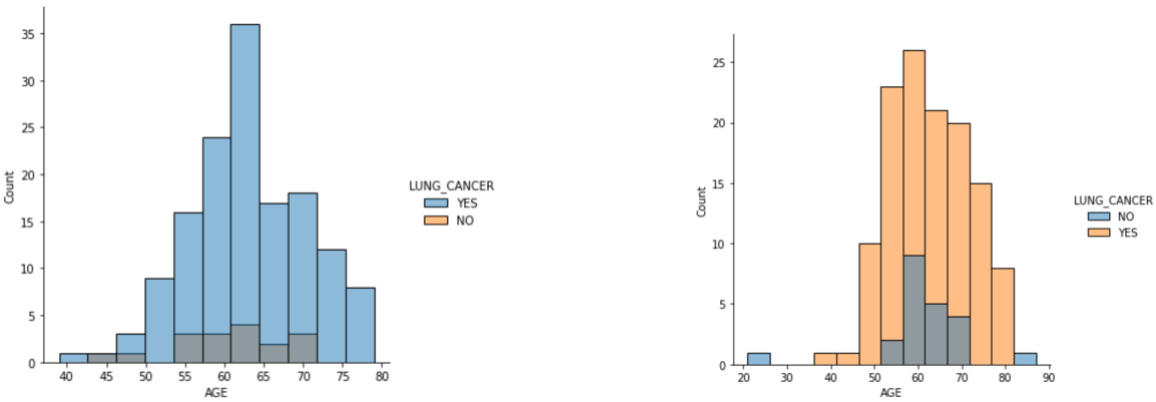


Figure 3. Participants were distributed according to age group and gender type in the balanced dataset.

A recurrence dispersion with infinite classes is shown via a histogram. It is an outline of a region comprised of square shapes with bases at the spans of the contrasting classes and regions proportionate to the frequencies of the contrasting classes. The square shapes are connected as the base fills up the areas between the class borders. The proportions of the squares that make up the statures correspond to the relative class frequencies and recurrence densities for various classes. Some key characteristics of the histograms are illustrated in Figure 4. The distribution of the dataset is shown in the histogram.

4. RESULTS AND ANALYSIS

Models capabilities, model predictions, analysis, and results are covered in this section with discussion. Figures 4, 5, 6 and 7 demonstrate the ML model's performance, exclusively for the lung cancer class, in terms of precision, recall, Accuracy and F1 score.

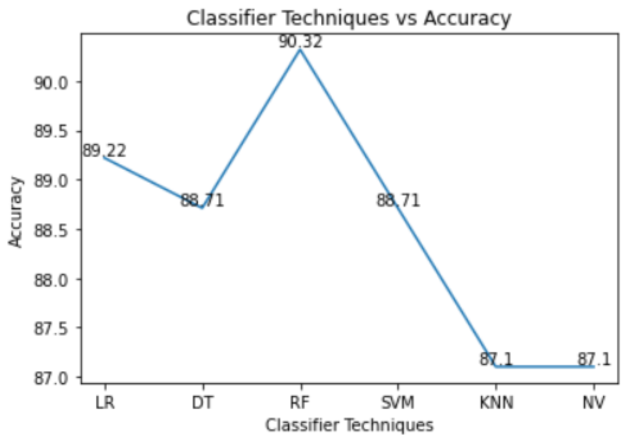


Figure 4. Performance Metrics for Accuracy

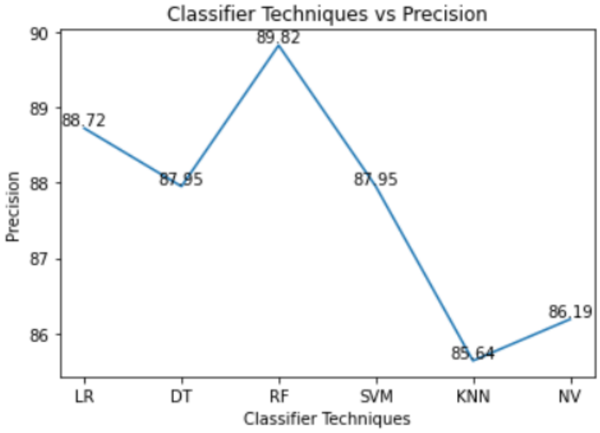


Figure 5. Performance Metrics for Precision

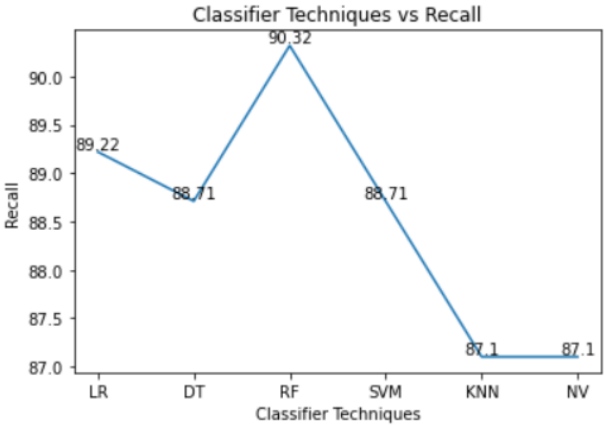


Figure 6. Performance Metrics for Recall

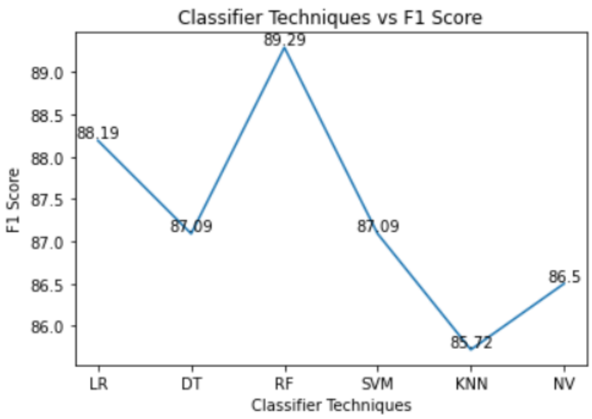


Figure 7. Performance Metrics for F1 Score.

Table 1 and the figure demonstrates unequivocally that the RF model is the most effective one among those models. Focusing on the performance metrics Accuracy, Precision, Recall and F1 Score the RF models have more performance which show that, with a high accuracy of 90.32%, precision of 89.82%, Recall 90.32% and F1 Score 89.29 can successfully identify the lung cancer from the non-lung cancer instances.

Table 1. Performance of ML models

	Accuracy	**Precision**	**Recall**	**F1 Score**
LR	89.22	88.72	89.22	88.19
DT	88.71	87.95	88.71	87.09
RF	90.32	89.82	90.32	89.29
SVM	88.71	87.95	88.71	87.09
KNN	87.1	85.64	87.1	85.72
NV	87.1	86.91	87.1	86.5

When we compare our results with others the copies of those found in past investigations are contrasted in Table 2 as shown that we got the best results in Random Forest.

Table 2. Proposed and Exiting Work Results

Matrix	**Exiting article [16]**	**Proposed Random Forest**
Accuracy	73.1%	90.32%
Precision	84%	89.82%
Recall	78.2%	90.32%
F1 Score	81%	87.09%

5. CONCLUSION

To conclude, factors can be extracted from massive databases of patient records by machine learning algorithms, which can then be used to forecast the likelihood of developing lung cancer with high accuracy. This might improve patient outcomes and save healthcare costs by allowing for an earlier diagnosis and better disease treatment. The study assessed how well different algorithms were able to foretell the development of lung cancer in individuals. The outcomes demonstrated that the random forest algorithm surpassed the competition in terms of better performance.

REFERENCES

[1] M. Guo *et al.*, "Autologous tumor cell–derived microparticle-based targeted chemotherapy in lung cancer patients with malignant pleural effusion," *Sci. Transl. Med.*, vol. 11, no. 474, Jan. 2019, doi: 10.1126/scitranslmed.aat5690.

[2] L. Zhang, Y. Hang, M. Liu, N. Li, and H. Cai, "First-Line Durvalumab Plus Platinum-Etoposide Versus Platinum-Etoposide for Extensive-Stage Small-Cell Lung Cancer: A Cost-Effectiveness Analysis," *Front. Oncol.*, vol. 10, Dec. 2020, doi: 10.3389/fonc.2020.602185.

[3] S. C, H. S A, and G. H L, "Artifact removal techniques for lung CT images in lung cancer detection," *Int. J. Data Informatics Intell. Comput.*, vol. 1, no. 1, pp. 21–29, Sep. 2022, doi: 10.59461/ijdiic.v1i1.14.

[4] E. Dritsas and M. Trigka, "Lung Cancer Risk Prediction with Machine Learning Models," *Big Data Cogn. Comput.*, vol. 6, no. 4, p. 139, Nov. 2022, doi: 10.3390/bdcc6040139.

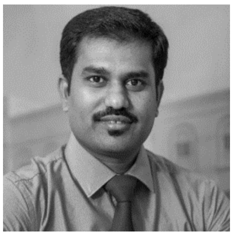
[5] Y. Zhang, B. Dai, M. Dong, H. Chen, and M. Zhou, "A Lung Cancer Detection and Recognition Method Combining Convolutional Neural Network and Morphological Features," in *2022 IEEE 5th International Conference on Computer and Communication Engineering Technology (CCET)*, IEEE, Aug. 2022, pp. 145–149. doi: 10.1109/CCET55412.2022.9906329.

[6] A. Aggarwal *et al.*, "The State of Lung Cancer Research: A Global Analysis," *J. Thorac. Oncol.*, vol. 11, no. 7, pp. 1040–1050, Jul. 2016, doi: 10.1016/j.jtho.2016.03.010.

[7] S. G. Spiro and G. A. Silvestri, "One Hundred Years of Lung Cancer," *Am. J. Respir. Crit. Care*

Med., vol. 172, no. 5, pp. 523–529, Sep. 2005, doi: 10.1164/rccm.200504-531OE.
[8] Y. She *et al.*, “Development and Validation of a Deep Learning Model for Non–Small Cell Lung Cancer Survival,” *JAMA Netw. Open*, vol. 3, no. 6, p. e205842, Jun. 2020, doi: 10.1001/jamanetworkopen.2020.5842.
[9] D. Deb, A. C. Moore, and U. B. Roy, “The 2021 Global Lung Cancer Therapy Landscape,” *J. Thorac. Oncol.*, vol. 17, no. 7, pp. 931–936, Jul. 2022, doi: 10.1016/j.jtho.2022.03.018.
[10] A. Hosny *et al.*, “Deep learning for lung cancer prognostication: A retrospective multi-cohort radiomics study,” *PLOS Med.*, vol. 15, no. 11, p. e1002711, Nov. 2018, doi: 10.1371/journal.pmed.1002711.
[11] V. K. Raghu *et al.*, “Validation of a Deep Learning–Based Model to Predict Lung Cancer Risk Using Chest Radiographs and Electronic Medical Record Data,” *JAMA Netw. Open*, vol. 5, no. 12, p. e2248793, Dec. 2022, doi: 10.1001/jamanetworkopen.2022.48793.
[12] K.-H. Yu *et al.*, “Reproducible Machine Learning Methods for Lung Cancer Detection Using Computed Tomography Images: Algorithm Development and Validation,” *J. Med. Internet Res.*, vol. 22, no. 8, p. e16709, Aug. 2020, doi: 10.2196/16709.
[13] P. Afshar *et al.*, “DRTOP deep learning-based radiomics for the time-to-event outcome prediction in lung cancer,” *Sci. Rep.*, vol. 10, no. 1, p. 12366, Jul. 2020, doi: 10.1038/s41598-020-69106-8.
[14] Y. Wu *et al.*, “Old age and *EGFR* mutation status in inoperable early-stage non-small cell lung cancer patients receiving stereotactic ablative radiotherapy: A single institute experience of 71 patients in Taiwan,” *Thorac. Cancer*, vol. 14, no. 7, pp. 654–661, Mar. 2023, doi: 10.1111/1759-7714.14786.
[15] S. Doppalapudi, R. G. Qiu, and Y. Badr, “Lung cancer survival period prediction and understanding: Deep learning approaches,” *Int. J. Med. Inform.*, vol. 148, p. 104371, Apr. 2021, doi: 10.1016/j.ijmedinf.2020.104371.
[16] Rajesh N., A. Irudayasamy, M. S. K. Mohideen, and C. P. Ranjith, “Classification of Vital Genetic Syndromes Associated With Diabetes Using ANN-Based CapsNet Approach,” *Int. J. e-Collaboration*, vol. 18, no. 3, pp. 1–18, Aug. 2022, doi: 10.4018/IJeC.307133.
[17] Y. Li, X. Wu, P. Yang, G. Jiang, and Y. Luo, “Machine Learning for Lung Cancer Diagnosis, Treatment, and Prognosis,” *Genomics. Proteomics Bioinformatics*, vol. 20, no. 5, pp. 850–866, Oct. 2022, doi: 10.1016/j.gpb.2022.11.003.
[18] C. Anil Kumar *et al.*, “Lung Cancer Prediction from Text Datasets Using Machine Learning,” *Biomed Res. Int.*, vol. 2022, pp. 1–10, Jul. 2022, doi: 10.1155/2022/6254177.
[19] E. Dritsas and M. Trigka, “Stroke Risk Prediction with Machine Learning Techniques,” *Sensors*, vol. 22, no. 13, p. 4670, Jun. 2022, doi: 10.3390/s22134670.
[20] T. Tazin, M. N. Alam, N. N. Dola, M. S. Bari, S. Bourouis, and M. Monirujjaman Khan, “Stroke Disease Detection and Prediction Using Robust Learning Approaches,” *J. Healthc. Eng.*, vol. 2021, pp. 1–12, Nov. 2021, doi: 10.1155/2021/7633381.
[21] G. Sailasya and G. L. A. Kumari, “Analyzing the Performance of Stroke Prediction using ML Classification Algorithms,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 6, 2021, doi: 10.14569/IJACSA.2021.0120662.

BIOGRAPHIES OF AUTHORS



Kumar Mohan is an experienced Lecturer from the University of Technology and Applied Sciences –Shinas, Oman. He has vast expeince of more than 17 years in teaching and research. He completed his Master of Computer Science from Sathyabama University in 2007 and completed his Bachelor of Computer Science and Engineering from Periyar University in 2003. He has received three funded projects from URG and MOHERI Oman. He has published many papers and has been a reviewer for many reputed journals. He can be contacted at email: kumar.mohan@shct.edu.om.

Bhraguram T M received his Doctorate degree from Kanpur University, India and his M.Sc. degree in Computer Science and Engineering (CSE) from Bharathidasan University, India. He has completed an M.E degree in Computer Science and Engineering (CSE) from VMKV Engineering College Tamilnadu, India and a Cyber security specialization from Govt. Law college, Mumbai. He has a Master of Business degree and is a holder in E-business. His distinguished career spans 15 years of academic and one year of corporate experience. He has published more than 19 articles, which include Scopus & Web of Science (WoS). He holds multifarious memberships from prominent professional bodies, specifically from IEEE, ACM, ISTE, IACSIT, CSTA, IAENGG, IAHFP and IARCP. He holds three patent registrations in various fields, and his research interests and areas are data mining, cloud-based services, digital forensics, Metaverse, and block chaining. He has won many awards and accolades during his career and is presently working as a lecturer at the University of Technology and Applied Sciences, Shinas, Sultanate of Oman. He can be contacted at email: bhraguram.thayyil@shct.edu.om.