BALTIC JOURNAL OF LAW & POLITICS
A Journal of Vytautas Magnus University
VOLUME 15, NUMBER 4 (2022)
ISSN 2029-0454

Cite: *Baltic Journal of Law & Politics* 15:4 (2022): 466-474
DOI:10.2478/bjlp-2022-004050

Comparative Analysis of Accuracy, Sensitivity & Specificity of Novel Decision Tree Algorithm to Maximise the Early Detection Rate of Lung Cancer in Comparison with Logistic Regression Algorithm

Y Kotesh kumar
Research Scholar, Department of Electronics and Communication Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamilnadu. India. Pincode: 602105.

R. Priyanka
Project Guide, Corresponding Author, Department of Electronics and Communication Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamilnadu. India. Pincode: 602105.

Received: August 8, 2022; reviews: 2; accepted: November 29, 2022.

Abstract

Aim: The objective of this research work is to maximise the early detection rate of lung cancer using the novel decision tree algorithm in comparison with the logistic regression algorithm. **Methods and Materials:** A total of 304 samples are collected from three lung cancer datasets available in kaggle. Group 1 represents the novel decision rate algorithm and group 2 represents the logistic regression algorithm.The G power calculation was done with 80 % of power and alpha of 0.05. **Results:** Novel decision tree has achieved the significance accuracy of 94.86 % compared to 80.11 %, by logistic regression algorithm. The novel decision tree algorithm has achieved the significance value of 0.044 when compared to the logistic regression algorithm. **Conclusion:** In this research, it was observed that the decision tree method outperforms the logistic regression approach in detecting lung cancer in the datasets that were considered.

Keywords

Novel Decision Tree Algorithm, Logistic Regression Algorithm, Lung Cancer, Image Processing, False detection, Machine Learning

INTRODUCTION

Lung cancer is the second most prominent cancer in men and women worldwide, and it is the leading cause of mortality in both men and women. Nowadays, cancer is a massive public health problem around the world. Therefore early detection and classification of lung cancer have become needed to cure the patients. (C. Subhash. 2013) Lung cancer begins to evolve while cells start to develop out of control. Most lung cancers start with inside the ducts that deliver air to the glands (ductal cancers). Some begin with inside the glands that make lungs expand (lobular cancers). A small variety of cancers start in different

tissues inside the lungs (A. Mariotto, E. Feuer. 2009, August). These cancers are called sarcomas and lymphomas and are not really thought of as lung cancers (S. I. Padma. 2019). Early identification of lung disease using image processing techniques is significant as it is connected with an expanded number of accessible treatment choices, expanded endurance, and improved personal satisfaction (G. Raghu. 2019). Innovative lung cancer detection CAD systems with deep learning and machine learning algorithms have applications in image recognition, video recognition, image classification, iris monitoring and medical image analysis (K. M. Chaitra. 2021) (Smerage et al. 2014) (Varma, Deekshitha Varma, and Priyanka 2022).

In recent years several articles have been published on lung cancer detection using novel decision tree algorithms. Many researchers have published in IEEE xplore around 102 research articles, around 5310 research articles in google scholar and around 640 research articles were published in Science Direct. Recently a lot of research has been done on a variety of Lung cancer using deep learning and machine learning for lung cancer detection (E. Williams. 2015). The growth and new technological innovations especially machine learning and artificial intelligence have a great impact on medical applications providing an effective supporting tool (P. E. Andersen. 2012) in different fields in medicine. Specifically with image processing (S. Thomas. 2009). For image processing segmentation and classification, different machine learning methods are proposed which support the radiologist with a clear opinion. In the survey different methodology and modification of pre-trained networks are used for segmentation, classification and analysis of affected MRI lung image. (Maiti et al. 2010) developed a method focused on optimised tumour detection and segmentation using decision trees. (Abinaya and Padma 2019) predicting lung cancer with deep learning. The dataset used is not up to the mark. In practice lung cancer is not easy to detect accurately within the early time and it is difficult to cure completely in later stages. They will weaken the lung and cause lung cancer (Hakim et al. 2015). In recent years, deep learning and machine learning algorithms have developed fast, and exceed human performance in visual tasks. (Sezer and Sezer 2020).

Previously our team has a rich experience in working on various research projects across multiple disciplines (Venu and Appavu 2021; Gudipaneni et al. 2020; Sivasamy, Venugopal, and Espinoza-González 2020; Sathish et al. 2020; Reddy et al. 2020; Sathish and Karthick 2020; Benin et al. 2020; Nalini, Selvaraj, and Kumar 2020).The existing systems use a smaller number of datasets which have major difficulty in detection of lung cancer. In order to overcome this, more datasets have been used in the proposed logistic regression algorithm which has been implemented using the novel decision tree algorithm. The accuracy is highly increased using the proposed innovative detection of lung cancer method using machine learning techniques.

MATERIALS AND METHODS

The research was conducted in the Department of Electronics and Communication Engineering, Saveetha School of Engineering at Saveetha Institute of Medical and Technical Sciences. This analysis consists of two different groups, a novel decision tree algorithm and a logistic regression algorithm. A total of 304 lung cancer samples were taken from two groups. The group 1 with 20 samples and group 2 with 20 samples were used in this analysis. The sample size was estimated for each group using a G power calculator with 80 % of pretest power, alpha error of 0.95, threshold value of 0.05, confidence level of 95 % (R, Sannasi, and Rajaguru 2019; Mutyalamma et al. 2017)

Decision Tree

Decision tree algorithm is a family of simple probabilistic classifiers based on applying decision theorem, from different vision objects and reducing the amount of data to be processed. It has been widely applied in various computer vision and image processing applications.

The steps involved in Decision tree algorithm are

Step 1: The images from the dataset are given as input to the decision tree algorithm.
Step 2: The images are processed using lung cancer reduction of gradient calculation
Step 3: Apply decision Tree on the subset of attributes that maximises classification of accuracy.
Step 4: Calculate accuracy of the classifier which measures the ability of the classifier to correctly classify the unknown sample.

Logistic Regression
The logistic regression is used in image processing for better accuracy results. Technically, it is a discrete differentiation operator, computing an approximation of the gradient of the image intensity function.
Step 1: Convert the true-colour RGB image to the grayscale lung cancer image.
Step 2: Pre-allocation of the filtered image matrix with zeros is done.
Step 3: Thresholding on the filtered image is carried out and the lung cancer-detected images are obtained.
The testing setup for the simulation of the decision tree and the logistic regression algorithm is done using the JUPYTER (2020) (Lavanya 2018). The basic system requirements for the JUPYTER software is 8th generation Intel i5 processor with 4GB RAM. It provides an interactive environment for algorithm development, data visualisation, data analysis, and numerical computation.

Statistical Analysis

IBM SPSS statistical tool was used for statistical analysis of this proposed work. The dependent and independent variables are analysed using independent sample t-tests. Independent sample t-tests were done for accuracy and mean deviation. Group 1 as logistic regression algorithm and group 2 as decision tree algorithm with 20 samples. Each group is considered for analysis of independent sample t tests. Accuracy is considered as a dependent variable whereas symmetry, compactness, correlation and energy are considered to be independent variables (Sathishkumar et al. 2019).

RESULTS

The novel lung cancer systems using decision tree give better accuracy and false identification rate when compared to the logistic regression algorithm.
Figure 1 represents the image of lung cancer patients with blood clots in their deep vein (pulmonary embolism) which is taken from the kaggle database for analysis. Fig. 2 shows the cancer developed in cells producing mucus substances which creates the larger impact in the lung.
Fig. 3 shows the statistical analysis of mean accuracy and groups. The mean accuracy and the standard deviation of the decision tree (85.80 %) is better than the logistic regression algorithm (75.10 %). X axis represents Decision tree algorithm vs logistic regression algorithm, Y axis represents Mean accuracy.
Table 1 shows the group statistics and accuracy of the decision tree and logistic regression algorithm with mean and standard deviation of 85.80 and 3.15, 75.10 and 3.38 respectively. The decision tree algorithm achieved a better standard error of 0.997 with the logistic regression algorithm.
Table 2 denotes the independent sample t- test of the group and mean accuracy of decision tree algorithm and logistic regression algorithm. There is a statistically significant difference in comparison with the decision tree and logistic regression algorithm ($p<0.05$). The decision tree algorithm has the highest accuracy (94.86 %) and the logistic regression algorithm has the lowest accuracy (80.11 %). Table 3 represents the practical values of accuracy for selected classifiers with a decision tree of 94.86% and Logistic regression of 80.11 %.

DISCUSSION

In this work it is observed that the decision tree algorithm has reduced the false detection rate when compared to the logistic regression algorithm by performing an independent sample t-test. The proposed algorithm reduces the false detection rate of lung cancer detection and improves the accuracy of detection.
According to the research, the decision tree algorithm appears to be superior to the logistic regression approach, and difficulties with false lung cancer identification are addressed by the decision tree algorithm employing a revolutionary lung cancer detection system. The decision tree method was suggested by (Chan et al. 2009) for higher accuracy and reducing the false detection rate, and it reached an accuracy of 88.52 percent. Using the gradients method, (Wang 2018) gave an overview of lung cancer detection strategies and obtained detection accuracy of 80.11 percent.
(Alloghani et al. 2019) proposed a decision tree that specified a lung cancer detection technique based on morphological edge detection and background differencing method with 86.89 % of accuracy achieved. (Bhandary et al. 2021; Odeh 2017) proposed a support vector model to identify the primary and secondary stages of lung cancer and achieved the detection accuracy of 95.5%. (Garud and Dhage 2021) used convolutional neural network and computed tomography images for detection of malignant or benign lung cancer stages with the accuracy of 87.4%. (Huber 2012) used the logistics regression algorithm with the computed tomography images to improve the detection rate of lung cancer and achieved the accuracy of 87.4%.
The proposed method focused on removal of false detection rate of lung cancer detection system by using two different detection algorithms. The limitation is that if the discontinuities persist then the identification rate of the system is affected. In the future work, a false detection rate of lung cancer removing approach based on the region techniques can be included for better identification ratio.

CONCLUSION

The novel decision tree algorithm for lung cancer system achieved the better detection rate in terms of accuracy. The decision tree algorithm produces significantly better accuracy (14 % higher) when compared to the logistic regression algorithm.

DECLARATIONS

Conflict of interests

No conflicts of interest in this Manuscript.

Author Contribution

Author YKK is involved in data collection, data analysis, and manuscript writing. Author RP is involved in conceptualization, guidance and critical review of manuscript.

Acknowledgements

The authors would like to express their gratitude towards Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (Formerly known as Saveetha University) for providing the necessary infrastructure to carry out this work successfully.

Funding

We thank the following organisations for providing financial support that enabled us to complete the study.

1. ABN info Sources Pvt Ltd, Chennai.
2. Saveetha University.
3. Saveetha Institute of Medical and Technical Sciences.
4. Saveetha school of Engineering.

REFERENCE

V. Krishnaiah, G. Narsimha, C. Subhash. 2013, December "Diagnosis of lung cancer prediction system using data mining classification techniques,"in International

Journal of Computer Science and Information Technologies, Vol. 4, issue 1, pp. 39-45.

J. J. Dignam, L. Huang, L. Ries, M. Reichman, A. Mariotto, E. Feuer. 2009, August. "Estimating cancer statistics and other-cause mortality in clinical trials and population-based cancer registry cohorts." Wiley InterScience[Online]. Vol. 115, issue 22.

Abinaya, M. P., and S. I. Padma. 2019. "A NOVEL SEGMENTATION AND CLASSIFICATION TECHNIQUES FOR MRI BRAIN TUMOUR IMAGES." ICTACT Journal on Image and Video Processing. https://doi.org/10.21917/ijivp.2019.0289.

Balekian, A. A., T. Henry, and G. Raghu. 2019. "Interstitial Lung Disease Is an Independent Cause of Lung Cancer Among Smokers." D19. LUNG CANCER SCREENING: NOVEL INSIGHTS ON PATIENT SELECTION AND OUTCOMES. https://doi.org/10.1164/ajrccm-conference.2019.199.1_meetingabstracts.a5891.

Bhandary, Abhir, Prabhu G. Ananth, Mustafa Basthikodi, and K. M. Chaitra. 2021. "Early Diagnosis of Lung Cancer Using Computer Aided Detection via Lung Segmentation Approach." International Journal of Engineering Trends and Technology. https://doi.org/10.14445/22315381/ijett-v69i5p213.

Brodrick, Emma, Antony Davies, Paul Neill, Louise Hanna, and E. Williams. 2015. "Detection of Lung Cancer Using Expired Breath Analysis by Ion Mobility Spectrometry." 11.1 Lung Cancer. https://doi.org/10.1183/13993003.congress-2015.pa2999.

Brünner, S., B. Langfeldt, and P. E. Andersen. 2012. Early Detection of Breast Cancer. Springer Science & Business Media.

Chan, Hiang Ping, Craig Lewis, and Paul S. Thomas. 2009. "Exhaled Breath Analysis: Novel Approach for Early Detection of Lung Cancer." Lung Cancer. https://doi.org/10.1016/j.lungcan.2008.05.020.

Dervisevic, Muamer, Maria Alba, Timothy E. Adams, Beatriz Prieto-Simon, and Nicolas H. Voelcker. 2021. "Electrochemical Immunosensor for Breast Cancer Biomarker Detection Using High-Density Silicon Microneedle Array." Biosensors & Bioelectronics 192 (July): 113496.

Devi, M. Shyamala, M. Shyamala Devi, A. N. Sruthi, and P. Balamurugan. 2017. "Artificial Neural Network Classification-Based Skin Cancer Detection." International Journal of Engineering & Technology. https://doi.org/10.14419/ijet.v7i1.1.10364.

Fazeli, Bahare, Pavel Poredos, Malay Patel, Peter Klein-Weigel, Mariella Catalano, Edwin Stephen, Mussaad M. Al Salman, et al. 2021. "Milestones in Thromboangiitis Obliterans. A Position Paper of the VAS-European Independent Foundation in Angiology/Vascular Medicine." International Angiology: A Journal of the International Union of Angiology, July. https://doi.org/10.23736/S0392-9590.21.04712-X.

Hakim, David N., Theo Pelly, Myutan Kulendran, and Jochem A. Caris. 2015. "Benign Tumours of the Bone: A Review." Journal of Bone Oncology. https://doi.org/10.1016/j.jbo.2015.02.001.

Kumarappa, S., and G. P. PrabhuKumar. 2004. "Experimental Investigation on a CNG Operated Stratified Scavenged Two Stroke SI Engine Using Embedded Systems." SAE Technical Paper Series. https://doi.org/10.4271/2004-28-0007.

Kuntamung, Kulrisa, Jaroon Jakmunee, and Kontad Ounnunkad. 2021. "A Label-Free Multiplex Electrochemical Biosensor for the Detection of Three Breast Cancer Biomarker Proteins Employing Dye/metal Ion-Loaded and Antibody-Conjugated Polyethyleneimine-Gold Nanoparticles." Journal of Materials Chemistry. B, Materials for Biology and Medicine, July. https://doi.org/10.1039/d1tb00940k.

Kurkure, Manasee, and Anuradha Thakare. 2016. "Introducing Automated System for Lung Cancer Detection Using Evolutionary Approach." International Journal Of Engineering And Computer Science. https://doi.org/10.18535/ijecs/v5i5.69.

Lavanya, S. 2018. "The Diagnosis of Lung Cancer Using Machine Learning." International Journal for Research in Applied Science and Engineering Technology. https://doi.org/10.22214/ijraset.2018.3490.

Maiti, B., M. N. Kundranda, T. P. Spiro, and H. A. Daw. 2010. "The Association of Metabolic

Syndrome with Triple-Negative Breast Cancer." Breast Cancer Research and Treatment. https://doi.org/10.1007/s10549-009-0591-y.

Mutyalamma, G., K. Komali, G. Pushpa, and Dadi Institute of Engineering and Technology. 2017. "Ranking and Fraud Review Detection for Mobile Apps Using KNN Algorithm." International Journal of Trend in Scientific Research and Development. https://doi.org/10.31142/ijtsrd8232.

R, Sannasi Chakravarthy S., Chakravarthy S. R. Sannasi, and Harikumar Rajaguru. 2019. "Lung Cancer Detection Using Probabilistic Neural Network with Modified Crow-Search Algorithm." Asian Pacific Journal of Cancer Prevention. https://doi.org/10.31557/apjcp.2019.20.7.2159.

Sathishkumar, R., K. Kalaiarasan, A. Prabhakaran, and M. Aravind. 2019. "Detection of Lung Cancer Using SVM Classifier and KNN Algorithm." 2019 IEEE International Conference on System, Computation, Automation and Networking (ICSCAN). https://doi.org/10.1109/icscan.2019.8878774.

Sezer, Aysun, and Hasan Basri Sezer. 2020. "Convolutional Neural Network Based Diagnosis of Bone Pathologies of Proximal Humerus." Neurocomputing. https://doi.org/10.1016/j.neucom.2018.11.115.

Smerage, Jeffrey B., William E. Barlow, Gabriel N. Hortobagyi, Eric P. Winer, Brian Leyland-Jones, Gordan Srkalovic, Sheela Tejwani, et al. 2014. "Circulating Tumor Cells and Response to Chemotherapy in Metastatic Breast Cancer: SWOG S0500." Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology 32 (31): 3483–89.

Supriya, M., and A. J. Deepa. 2020. "A Novel Approach for Breast Cancer Prediction Using Optimized ANN Classifier Based on Big Data Environment." Health Care Management Science 23 (3): 414–26.

Wang, Lingzhi. 2018. "Novel Exosome Proteins as Potential Biomarkers for Early Detection of Lung Cancer." Journal of Cancer Diagnosis. https://doi.org/10.4172/2476-2253-c1-001.

Varma, D. R. Deekshitha, D. R. Deekshitha Varma, and R. Priyanka. 2022. "Performance Analysis of Novel Iris Monitoring System Based on Canny Detection Algorithm in Comparison with Prewitt Algorithm." 2022 2nd International Conference on Innovative Practices in Technology and Management (ICIPTM). https://doi.org/10.1109/iciptm54933.2022.9754088.

Figures and Tables

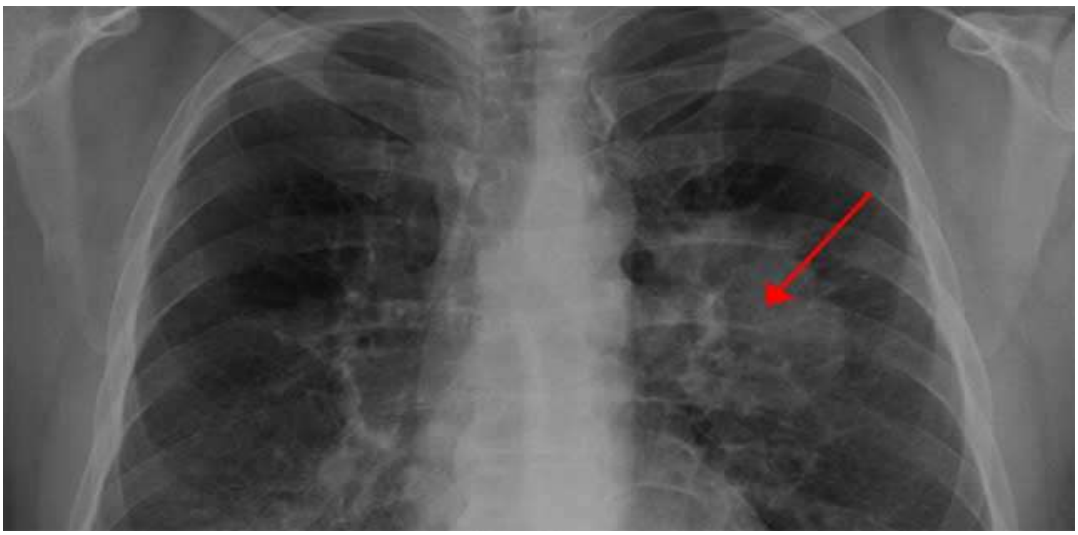

Fig. 1. The image shows lung cancer of the patient with blood clot in their deep vein (pulmonary embolism) taken from the kaggle database.

Fig. 2. Cancer develops in cells that produce mucus substances which denote the larger impact in the lung and the image was taken from the kaggle lung cancer dataset.

Fig. 3. Statistical analysis of mean accuracy and groups. The mean accuracy and the standard deviation of Decision Tree (85.80 %) is better with the Logistic regression algorithm (75.10 %). X axis: Decision Tree algorithm vs Logistic regression algorithm, Y axis: Mean Accuracy ± 2 SE.

Table 1. Comparison of the group statistics and accuracy of decision tree and logistic regression algorithm with mean and standard deviation of 85.80 and 3.15, 75.10 and 3.38 respectively. The decision tree algorithm achieved a better standard error of 0.997 when compared to the logistic regression algorithm. There is a statistically significant difference in accuracy of the 2 algorithms using independent sample t-tests

Group Statistics					
	Classifier	**N**	**Mean**	**Std. Deviation**	**Std. Error Mean**
ACCURACY	**Decision Tree**	20	85.800	3.15524	.99778
	Logistic regression	20	75.100	3.38132	1.06927

Table 2. Independent sample t-test analysis for group and accuracy of decision tree algorithm and logistic regression algorithm. There is a statistically significant difference in comparison with the decision tree and logistic regression algorithm with $p<0.05$ using t-test. The decision tree algorithm had the highest accuracy (94.86 %) and the logistic regression algorithm the lowest accuracy (80.11 %).

Independent samples test										
		Levene' Test For equality of Variances		**T-test For Equality of Means**						
		F	**sig**	**t**	**df**	**Sig 2**	**Mean difference**	**Std error difference**	**95% confidence interval of the difference**	
									lower	**upper**
Accuracy	**Decision Tree**	.390	.04040	7.316	18	<.001	10.70000	1.46249	7.62741	13.77259
	Logistic Regression			7.316	17.914	.001	10.70000	1.46249	7.62636	13.77364

Table 3. The Practical values of accuracy for selected classifier with decision tree of 94.86 % and logistic regression of 80.11 %

Algorithm	Accuracy (%)
Decision Tree	94.86 %
Logistic Regression	80.11 %