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): 252-259
DOI: 10.2478/bjlp-2022-004026

Analysis of Lung Cancer Detection System for Better Identification Rate using Novel Support Vector Machine Algorithm 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: This study's objective is to determine whether the novel Support Vector Machine (SVM) method can lower the false detection rate of lung cancer when compared to the logistic regression approach. **Materials and Methods**: Three lung cancer datasets accessible from kaggle and yielded a total of 304 samples. The SVM method is represented as group 1, while the logistic regression algorithm is represented by group 2. The G power calculation was done with an alpha of 0.05 and a power of 80%. **Results:** The accuracy of the SVM algorithm is 93.59% better than the 80.11% for the logistic regression approach. The SVM and logistics regression algorithm has the significance of (p=.0534). **Conclusion**: The SVM algorithm is proven to be much more accurate in this work when compared to the logistic regression algorithm.

Keywords

Novel Support Vector Machine (SVM), Logistic Regression Algorithm, Lung Cancer, Image Processing, False detection, Machine Learning, Medical application

INTRODUCTION

Lung cancer is one among the foremost dangerous and main reasons for early death of the person across the world (Santosh et al. 2019). 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 in medical applications (Brünner, Langfeldt, and Andersen 2012). Lung Cancer begins to evolve while cells start to develop out of control. Most lung cancer starts with inside the ducts that deliver air to the glands . Lung cancer is the major major cause of death worldwide, affecting a large number of people. For huge datasets, machine learning techniques are applied in hospitals and medical applications (Raja et al. 2020; Lavanya 2018). In recent years several articles have been published on Lung cancer detection using support vector machine algorithms with Logistic regression classifiers as well as for iris monitoring (Vo 2020) (Varma,

Deekshitha Varma, and Priyanka 2022). Many researchers have published in IEEE xplore around 212 research articles, around 6311 research articles in Google Scholar and around 840 research articles were published in science direct. Recently a lot of research has been done on a variety of lung cancer using machine learning and deep learning for lung cancer detection. (Brodrick et al. 2015) According to a survey, many research projects involve the detection and care of lung cancer using subset methods and support vector machine algorithms and achieve a better detection rate. (Ubaldi et al. 2021) proposed noise reductions for obtaining better features in the dataset with the wrapper method achieved better accuracy. (Maiti et al. 2010) developed a method focused on optimised detection and Segmentation of lung cancer image using K-Nearest Neighbour algorithm. (Balekian, Henry, and Raghu 2019; Davies 2012) presented a more accurate method for using Logistic Regression classifiers to identify Lung Cancer in Images. (Juiling et al. 2020) A support vector machine technique using a repetition thresholding approach has been suggested, and it offers the best accuracy of 80% for various datasets.

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 current approach has the flaw of not selecting the best attributes from the photographs. As a result, the purpose of this research is to pick relevant features using the SVM method rather than logistic regression to minimise false detection rates and improve apparent accuracy.

MATERIALS AND METHODS

The study was carried out at the Saveetha Institute of Medical and Technical Sciences' Department of Electronics and Communication Engineering. SVM algorithm and logistic regression method are the two different groups used in this research. 304 lung cancer samples total were collected from two groups. In this research, the group 1 with 20 samples and the group 2 with 20 samples were both used. Using a G Power calculator with an 80 percent pretest power, alpha error of 0.95, threshold value of 0.05, and a 95 percent confidence level, the sample size was determined for each group (Wu and Zhao 2017).

Support Vector Machine

Support Vector Machine is a tool for obtaining meaningful structural information from multiple versions of visual objects in order to decrease the quantity of data to be processed. It has seen widespread use in a variety of computer vision applications.

The Support Vector Machine algorithm's stages are
Step 1: The SVM algorithm is fed photos from the dataset as input.
Step 2: A lung cancer divergence reduction computation is used to process the images.
Step 3: SVM should be used on the subset of characteristics with the highest classification accuracy.
Step 4: Calculate the classifier's accuracy, which evaluates the classifier's ability to accurately categorise the unknown sample.

Logistic Regression

The Supervised Learning methodology includes the popular machine learning technique of logistic regression. It is used to forecast the categorical dependent variable based on a set of independent factors.

Step 1: Logistic Regression has been fitted to the Training dataset.
Step 2: Pre-processing of dataset
Step 3: Predicting the outcome of a test.

Step 3: Visualising the test set outcome

The testing setup was carried out with the help of JUPYTER software (Navlani, Fandango, and Idris 2021). The tests are run on a device equipped with an 8th generation Intel i5 processor with 8GB RAM. For the creation of algorithms, visualization of data, analysis techniques, and numerical calculation, it offers an interactive environment.

Statistical Analysis

Lung cancer detection systems assessment is performed using SPSS (Brünner, Langfeldt, and Andersen 2012). For SVM and Logistic Regression, mean, standard deviation, and standard error are determined in SPSS. Using independent samples t - test, the dependent and independent variables are compared. Group 1 as Logistic regression algorithm and group 2 as support vector machine algorithm with 20 samples each groups are considered for analysis of independent sample t test. Accuracy is considered as a dependent variable whereas entropy, Correlation and energy are regarded as separate factors.

RESULTS

When compared to the Logistic Regression Algorithm, the innovative support vector machine systems for lung cancer provide superior accuracy and a lower false identification rate.

Figure 1 The image was derived from the Kaggle dataset and depicts the scanning of the right side spiral CT scan data used to identify early lung cancer in patients. Figure 2 represents the affected part of the right side lung due to pathologically malignant or other types of disease which are taken to form the dataset. Figure 3 shows the independent sample Test with the mean accuracy of the Novel Support vector machine (91.50%) is better than the Logistic regression algorithm (75.10%). And the acceptable error with Mean Accuracy ± 2 SD.

Table 1 Analyzing the group statistics for Novel SVM and Logistic Regression accuracy, with mean and SD of 91.50 and 4.9497, 75.10 and 3.38132, respectively. When compared to the Logistic Regression approach, the Novel SVM algorithm's standard error was superior, ending in at 1.069. Table 2 An independent sample t-test was used to compare the novel SVM algorithm with the logistic regression technique, and the results showed a statistically significant difference between the two when P 0.05 was used. Table 3 represents the Practical values of accuracy for selected classifier with SVM of 93.59% and Logistic regression of 80.11%

DISCUSSION

Through the use of an independent sample t-test, it has been determined in this study that the innovative support vector machine method has a lower false detection rate than the logistic regression technique.The suggested approach lowers the percentage of false detection rate of lung cancer detection and improves the accuracy of detection which is suitable for medical applications.

According to the analysis, SVM algorithm appears to be superior than logistic regression algorithm, and problems with false lung cancer identification are resolved by SVM algorithm utilising new lung cancer detection methods (Balekian, Henry, and Raghu 2019; Chan, Lewis, and Thomas 2009). proposed the SVM algorithm for better accuracy and removing the false detection rate and achieved an accuracy of 93.59% (Wang 2018). employing the gradients methodology, he offered a summary of lung cancer detection methods and attained detection results of 78.10 percent (Kurkure and Thakare 2016). proposed a SVM that specified a lung cancer morphological edge detection and ambient differencing-based detection method with 84.2% of accuracy achieved. (Bhandary et al. 2021) proposed a logistic regression algorithm for Fuzzy logic controller-based simulation of the maximum power point tracking (MPPT) technique that achieves a 78 percent

accuracy (Chan, Lewis, and Thomas 2009) suggested a sobel method for ideal threshold values of digital pictures in edge detection with 80% accuracy (Chan, Lewis, and Thomas 2009; Brodrick et al. 2015) proposed a SVM technique scheduling for genuine and semi datasets with 79.5 percent accuracy attained.

The proposed algorithm calculates the optimal features for improving the identification rate. The limitation of the systems is when the intensity of the image decreases the identification rate of the system also reduces. In future, it can be preprocessed to improve the intensity to achieve a better identification rate and which is suitable for medical applications.

CONCLUSION

Based on the obtained results the Support Vector Machine (SVM) Technique based on machine learning performs 93.59% accuracy (13% higher) compared to the Logistic regression technique yields an accuracy of 80.11 percent when employing the suggested innovative lung cancer detection method.

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 research.

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

REFERENCE

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.

Benin, S. R., S. Kannan, Renjin J. Bright, and A. Jacob Moses. 2020. "A Review on Mechanical Characterization of Polymer Matrix Composites & Its Effects Reinforced with Various Natural Fibres." *Materials Today: Proceedings* 33 (January): 798–805.

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.

Davies, E. R. 2012. *Computer and Machine Vision: Theory, Algorithms, Practicalities*. Academic Press.

Gudipaneni, Ravi Kumar, Mohammad Khursheed Alam, Santosh R. Patil, and Mohmed Isaqali Karobari. 2020. "Measurement of the Maximum Occlusal Bite Force and Its Relation to the Caries Spectrum of First Permanent Molars in Early Permanent Dentition." *The Journal of Clinical Pediatric Dentistry* 44 (6): 423–28.

Juiling, Suzika, Universiti Malaysia Sabah, Scholena Kinit Leon, Jamirus Jumian, Sandy Tsen, Yew Leung Lee, Eyen Khoo, et al. 2020. "Conservation Assessment and Spatial Distribution of Endemic Orchids in Sabah, Borneo." *Nature Conservation Research*. https://doi.org/10.24189/ncr.2020.053.

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.

Nalini, Devarajan, Jayaraman Selvaraj, and Ganesan Senthil Kumar. 2020. "Herbal Nutraceuticals: Safe and Potent Therapeutics to Battle Tumor Hypoxia." *Journal of Cancer Research and Clinical Oncology* 146 (1): 1–18.

Navlani, Avinash, Armando Fandango, and Ivan Idris. 2021. *Python Data Analysis: Perform Data Collection, Data Processing, Wrangling, Visualization, and Model Building Using Python*. Packt Publishing Ltd.

Raja, Rohit, Sandeep Kumar, Shilpa Rani, and K. Ramya Laxmi. 2020. *Artificial Intelligence and Machine Learning in 2D/3D Medical Image Processing*. CRC Press.

Reddy, Poornima, Jogikalmat Krithikadatta, Valarmathi Srinivasan, Sandhya Raghu, and Natanasabapathy Velumurugan. 2020. "Dental Caries Profile and Associated Risk Factors Among Adolescent School Children in an Urban South-Indian City." *Oral Health & Preventive Dentistry* 18 (1): 379–86.

Santosh, K. C., Sameer Antani, D. S. Guru, and Nilanjan Dey. 2019. "Medical Imaging." https://doi.org/10.1201/9780429029417.

Sathish, T., and S. Karthick. 2020. "Gravity Die Casting Based Analysis of Aluminum Alloy with AC4B Nano-Composite." *Materials Today: Proceedings* 33 (January): 2555–58.

Sathish, T., D. Bala Subramanian, R. Saravanan, and V. Dhinakaran. 2020. "Experimental Investigation of Temperature Variation on Flat Plate Collector by Using Silicon Carbide as a Nanofluid." In *PROCEEDINGS OF INTERNATIONAL CONFERENCE ON RECENT TRENDS IN MECHANICAL AND MATERIALS ENGINEERING: ICRTMME 2019*. AIP Publishing. https://doi.org/10.1063/5.0024965.

Sivasamy, Ramesh, Potu Venugopal, and Rodrigo Espinoza-González. 2020. "Structure, Electronic Structure, Optical and Magnetic Studies of Double Perovskite Gd2MnFeO6 Nanoparticles: First Principle and Experimental Studies." *Materials Today Communications* 25 (December): 101603.

Ubaldi, L., V. Valenti, R. F. Borgese, G. Collura, M. E. Fantacci, G. Ferrera, G. Iacoviello, et al. 2021. "Strategies to Develop Radiomics and Machine Learning Models for Lung Cancer Stage and Histology Prediction Using Small Data Samples." *Physica Medica: PM: An International Journal Devoted to the Applications of Physics to Medicine and Biology: Official Journal of the Italian Association of Biomedical Physics* 90 (September): 13–22.

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.

Venu, Harish, and Prabhu Appavu. 2021. "Experimental Studies on the Influence of Zirconium Nanoparticle on Biodiesel–diesel Fuel Blend in CI Engine." *International Journal of Ambient Energy* 42 (14): 1588–94.

Vo, Jessica. 2020. *Using Supervised Machine Learning for Early Lung Cancer Detection*.

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.

Wu, Qing, and Wenbing Zhao. 2017. "Small-Cell Lung Cancer Detection Using a Supervised Machine Learning Algorithm." *2017 International Symposium on Computer Science and Intelligent Controls (ISCSIC)*. https://doi.org/10.1109/iscsic.2017.22.

FIGURES AND TABLES

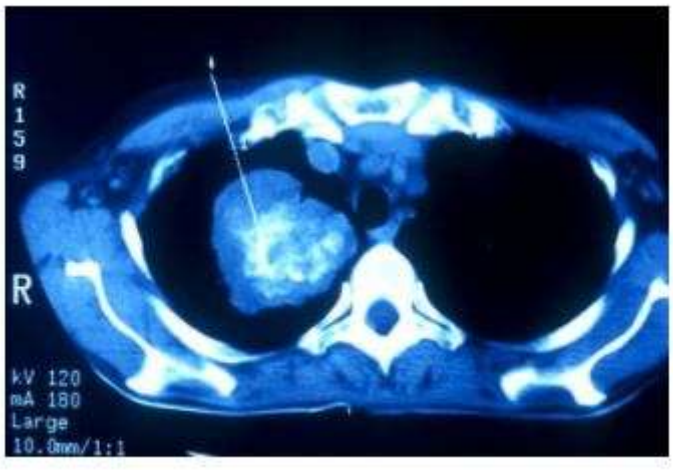

Fig. 1. The image was derived from the Kaggle dataset and depicts the scanning of the right side spiral CT scan data used to identify early lung cancer in patients.

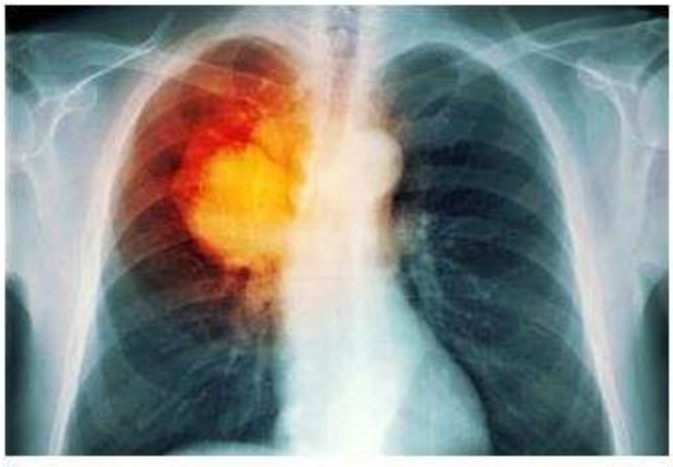

Fig. 2. The affected part of the lung on the left side, due to pathologically malignant or other types of disease which are taken from the kaggle dataset.

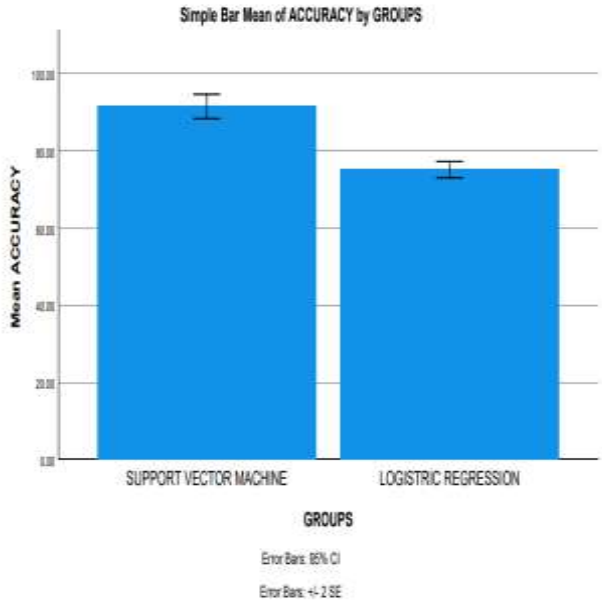

Fig. 3. The independent sample Test with the mean accuracy of the Novel Support vector machine (91.50%) is better than the Logistic regression algorithm (75.10%). And the acceptable error with Mean Accuracy ± 2 SD.

Table 1. Analyzing the group statistics for Novel SVM and Logistic Regression accuracy, with mean and SD of 91.50 and 4.9497, 75.10 and 3.38132, respectively. When compared to the Logistic Regression approach, the Novel SVM algorithm's standard error was superior, ending in at 1.069.

Group Statistics					
	Classifier	**N**	**Mean**	**Std. Deviation**	**Std. Error Mean**
ACCURACY	**SVM**	20	91.50	4.9497	1.0692
	Logistic regression	20	75.10	3.38132	1.5652

Table 2. An independent sample t-test was used to compare the novel SVM algorithm with the logistic regression technique, and the results showed a statistically significant difference between the two when P 0.05 was used.

Independent samples test										
		Levene' Test For equality of Variances		**T-test For Equality ofMeans**						
		F	**sig**	**t**	**df**	**Sig 2**	**Mean difference**	**Std error difference**	**95% confidence interval of the difference**	
									lower	**upper**
Accuracy	**SVM**	.402	.0534	8.652	18	<.001	6.20000	1.8956	12.417	20.38253
	Logistic Regression			8.652	15.898	<.001	6.20000	1.8956	12.379	20.42061

Table 3: The Practical values of accuracy for selected classifier with SVM of 93.59% and Logistic regression of 80.11%

Algorithm	**Accuracy(%)**
SVM	93.59%
LOGISTIC REGRESSION	80.11%