IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 1, October 2025
Impact Factor: 7.67

Lung Cancer Prediction Using Machine Learning: A Comparative Analysis of KNN, SVM, Random Forest, and Logistic Regression

Emmanuel Ifeoluwa Oyerinde, Abosede Ibironke Ojo, Adedoyin Samuel Adebanjo, Aisha Omorinbola Ajao, Afoluwajuwonlo Obaoye, Opeyemi Adelowo, Royce Nwoko, Emmanuel Alexander, Chidalu Chukwudebelu
Department of Information Technology
Babcock University, Nigeria
oyerindee@gmail.com

Abstract: *Among the major issues of cancer-associated fatalities universally is lung cancer, and survival rates are heavily reliant on prompt and precise diagnosis. Conventional diagnostic techniques, while effective, frequently miss early-stage lung cancer detection. By examining intricate patterns in medical data, machine learning offers a reassuring method for enhancing the prediction of lung cancer. However, the algorithm and optimization strategies employed determine how successful machine learning models are. This study explores the comparative evaluation of four machine learning approaches for the prediction of lung cancer: Random Forest, Logistic Regression, K-Nearest Neighbors (KNN), and Support Vector Machine (SVM). To improve model performance, the Kaggle dataset was preprocessed, encoded, and put through feature selection procedures. Hyperparameter tuning was used to refine model parameters acceptable to upsurge accuracy still more. Key performance pointers counting accuracy, precision, recall, and F1-score were accustomed to evaluate the models. The findings demonstrate with an accuracy of 90%, the Logistic Regression method performed best, with other models exhibiting varied degrees of performance. The outcome of this work highlights the significance of model assortment and parameter optimization, as well as the promise of machine learning in the prediction of lung cancer. Future research could explore deep learning approaches and integrate additional patient data to enhance predictive performance. Ultimately, leveraging machine learning for lung cancer diagnosis could lead to earlier detection, better persevering consequences, and a significant decrease in death rates.*

Keywords*:* Lung cancer, Machine learning, Random Forest, Logistic Regression, Hyperparameter tuning

DOI: 10.48175/IJARSCT-29107