# *ABSTRACT*

*Lung cancer remains to be a major cause of cancer-related mortality worldwide, that underscores the necessity of an early and precise diagnosis. This study investigates how feature selection can improve machine learning models for the diagnosis of lung cancer. Utilizing the Lung Cancer Dataset, we employed ranking-based feature selection methods, specifically Information Gain and Gini Decrease, for determining the most important features for model training. Three classifiers Random Forest, Decision Tree, and Support Vector Machine (SVM) were evaluated using measures including AUC, Classification Accuracy, F1 Score, Precision, Recall, and MCC. The results demonstrated notable improvements in model performance with feature selection, highlighting its importance in reducing computational complexity and improving diagnostic accuracy. The Random Forest model demonstrated the most favorable AUC performance of 0.980 and an accuracy of 94.4%, underscoring its robustness in predicting lung cancer outcomes. This study emphasizes the value of integrating feature selection into machine learning workflows to build efficient, interpretable, and clinically applicable models for lung cancer diagnosis.*

*Keywords: Feature Selection, Lung cancer diagnosis, Machine learning, Random Forest*