Adenox: A Machine Learning Model to Predict Molecular Targets in Non-Small Cell Lung Cancer

Melike Cömert[1], Morteza Faulady[1], Özge Pasin[2], Melih Şimşek[3]

[1]Bezmialem Vakif University, Faculty of Medicine, Istanbul, Türkiye

[2]University of Health Sciences, Faculty of Medicine, Department of Biostatistics and Medical Informatics, Istanbul, Türkiye

[3]Medipol Health Group, Department of Internal Medicine, Division of Medical Oncology, Istanbul, Türkiye

Introduction: Targeted therapies significantly improve survival in non-small cell lung cancer (NSCLC), one of the deadliest cancers. Identifying molecular targets is crucial for determining patient eligibility for targeted therapies; however, the lengthy process of obtaining results delays treatment initiation. This study aims to develop *Adenox*, a clinically practical machine learning model, to predict molecular targets using clinical features before test results are available.

Methods: This retrospective study included 281 NSCLC patients, comprising 80(28.5%) with EGFR, 17(6%) with ALK, 13(4.6%) with KRAS, 3(1.1%) with ROS1 mutations, 107(38.1%) with PD-L1 positivity and 85(30.2%) without molecular targets. Clinical features were categorized into demographics, radiological features, laboratory findings and pathological findings. Significant features identified via Logistic Regression were used to build the final model with Random Forest.

Results: Associated factors with targets included EGFR: gender, smoking, pack-years, alcohol consumption, histology, brain/bone metastases ($p<0.001$), asthma ($p=0.008$), tumor size ($p=0.006$), and spiculation ($p=0.035$); ALK: histology ($p=0.049$), LDH levels ($p=0.043$); PD-L1: smoking, pack-years ($p<0.001$), alcohol consumption ($p=0.002$), brain metastases ($p=0.003$), and histology ($p=0.004$). The most significant independent prognostic factors for EGFR were being female, not smoking, brain/bone metastases, spiculation, and pneumonic infiltration; for ALK, pleural metastasis; and for PD-L1, smoking. Receiver Operating Characteristic (ROC) analysis showed predictive performance for EGFR (AUC=0.887), PD-L1 (AUC=0.679), ALK (AUC=0.798), KRAS (AUC=0.674), and ROS1 (AUC=0.911).

Discussion: By demonstrating clinical effectiveness, our model highlights the potential of artificial intelligence to assist clinicians. The model can predict which targets may be associated with patients with specific characteristics. This approach can help determine the targets to analyze, particularly in patients with limited tumor tissue. We aim to integrate it into routine practice, hoping to improve survival and quality of life.

Keywords: Non-small cell lung cancer, prediction, molecular targets, machine learning