

MACHINE LEARNING DRIVEN IN SILICO SCREENING IN THE KNIME PLATFORM AND MOLECULAR DOCKING TO IDENTIFY PHYTOCHEMICALS WITH ALK INHIBITORY ACTIVITY

P.T. Dilshan¹, L.R.L.S. Kumari² and W.R.P. Wijesinghe^{2*}

¹*Department of Molecular Biology and Biotechnology, University of Peradeniya, Peradeniya, Sri Lanka.*

²*Department of Botany, University of Peradeniya, Peradeniya, Sri Lanka.*

*priyangaw@sci.pdn.ac.lk

ALK⁺ (Anaplastic lymphoma kinase) lung cancer, a subtype of non-small cell lung cancer, occurs due to aberrant activation of the ALK gene, commonly due to its fusion with other genes. Despite the presence of approved drugs against ALK, emerging resistance necessitates the discovery of safer and effective alternatives. Traditional drug discovery is a time-consuming and resource-intensive process. Consequently, this study aimed to identify potential ALK inhibitors from phytochemicals based on quantitative structure-activity relationship via Machine learning (ML) in the KNIME platform, followed by assessment of drug-likeness, molecular docking, and toxicity predictions. Using a modified TeachOpenCADD workflow, three ML models, namely Random Forest, Artificial Neural Network and Support Vector Machine, were trained to predict phytochemicals exhibiting structural similarity to known ALK inhibitors. Subsequently, the oral drug-likeness of predicted molecules was assessed via Lipinski's rule of five. Filtered compounds were subjected to molecular docking analysis using PYRX software employing the ALK kinase domain (PDB ID: 2XP2) as the macromolecule. Finally, toxicity prediction was performed using ProTox-3.0. ML models predicted ten phytochemicals with potential ALK inhibitory activity, achieving over 97% accuracy. All candidate molecules complied with three or more of Lipinski's criteria for drug likeness. Molecular docking revealed nine molecules with binding affinity less than the inclusive binding affinity threshold of -5 kcal mol^{-1} and six molecules exhibited a lower binding affinity than the stringent threshold of -7 kcal mol^{-1} . Based on toxicity predictions, out of six, only four phytochemicals, namely, Luteolin, Mearnsetin, Lichochalcone D, and Murayaquinone A, exhibited lower toxicity than the positive control, Crizotinib. Myricetin and Fisetin, with strong binding affinities, are worth exploring despite their toxicity, provided that the effective dose can be achieved within the therapeutic window. These findings emphasised the importance of phytochemicals as promising candidates for the development of next-generation inhibitors for ALK-positive lung cancers despite emerging resistance.

Keywords: CADD, Lipinski's rule of five, Lung cancer, PYRX, Tyrosine kinase inhibitors