Abstract No: 98 ***Life Sciences***

KNIME AND MACHINE LEARNING-BASED PREDICTION OF PDGFRA INHIBITORS FROM PHYTOCHEMICALS AS POTENTIAL CANCER DRUGS

B. Swetha[1], L.R.L.S. Kumari[2] and W.R.P. Wijesinghe[2,3*]

[1]Department of Molecular Biology and Biotechnology, Faculty of Science, University of Peradeniya, Peradeniya, Sri Lanka.
[2]Department of Botany, Faculty of Science, University of Peradeniya, Peradeniya, Sri Lanka.
[3]Postgraduate Institute of Science, University of Peradeniya, Peradeniya, Sri Lanka.
**priyangaw@sci.pdn.ac.lk*

Cancer is a major worldwide concern and is a heterogenic disease that involves various genes. By inducing cell division and angiogenesis, the Platelet-Derived Growth Factor Receptor Alpha (PDGFRA) gene plays a significant role in the development and progression of cancers like gastrointestinal stromal tumours and medulloblastoma. Tumor development can be inhibited by specifically blocking the activity of mutant PDGFRA using PDGFRA inhibitors. A PDGFRA inhibitor drug discovery approach is high-throughput screening, which is also time-, labour-, and cost-intensive. Thus, Quantitative Structure-Activity Relationship (QSAR), a computational modeling-based technique to determine the structural and biological similarities between chemical compounds, is highly recommended for drug discovery since it is productive and cost-efficient. Potential cancer drugs can be effectively detected with the combination of Machine Learning (ML) and the QSAR technique. In this study, KNIME v4.7.0 was used along with cheminformatics extensions to detect phytochemicals similar to PDGFRA inhibitors by using TeachOpenCADD KNIME. Simplified Molecular Input Line Entry System (SMILES) notations of 394 molecules were retrieved from the ChEMBL database. These molecules consist of 18 PDGFRA inhibitors, eight rheumatoid arthritis medications, and phytochemicals, including the phytochemicals common in Sri Lankan flora. The MACCS (Molecular Access System) fingerprint of molecules was calculated to perform the ML-based modelling to distinguish the active drugs against PDGFRA. Subsequently, the drug-likeliness of predicted compounds was evaluated by Lipinski's rule of five. The accuracy of Artificial Neural Networks (ANN), Random Forest (RF), and Support Vector Machine (SVM) ML models were 91.86%, 92.88%, and 92.2% respectively. The RF model predicted Kaempferol-3-O-P-D-glucoside while Kaempferol-3-O-P-D-glucoside and Kazinol J were predicted as potential PDGFRA inhibitors in ANN. The results of the SVM model included known PDGFRA inhibitors only. Therefore, the ANN model is efficient in drug discovery. However, further *in-vitro* and *in-vivo* experiments will enhance the reliability of using these predicted phytochemicals as cancer drugs.

Keywords: Cancer drugs, KNIME, Machine learning, PDGFRA inhibitors, Phytochemicals