BIN-O2

A Novel approach based on generative deep learning for inverse QSAR modeling

Abdelbasset Boukelia*, Khalid Bouhedjar

Unité Bioinformatique, Centre de Recherche en Biotechnologie, Constantine, Algérie.

*a.boukelia@crbt.dz

Abstract

Lung cancer is one of the worldwide deadly cancers and the second most common cancer in men and women. Although chemotherapy plays a key role in lung cancer treatment, resistance and serious adverse effects on normal tissue are the major causes of clinical cancer chemotherapy failure. In the recent decades, due to the evolution of artificial intelligence, especially machine learning & deep learning methods, drug discovery and medicine are now benefiting from these models through the automation of new chemical entities' inventions and the mining of large compound databases. In this work, we propose novel fingerprint approach based on variation alautoencoder and transfer learning to generate new compounds with the targeted activity. Our proposed pipeline consists of three main steps. First, we processed the textual compound data by calculating the fingerprints of each compound. Then, we have adapted and trained a variation alautoenoder model to generate compounds variants of the desired activity. Finally, we have built a new model by using transfer learning from the trained variation alautoencoder to a new MLP model for predicting IGC50. With this combination, we were able to achieve a determination coefficient up to 80,12%. In addition, we have generated in silico compounds for the targeted activity. The results of our experiments have been validated by analyzing molecules from a dataset LNS-NCI-H522.

Keywords: Drug Design, Quantitative Structure-Activity, Relationship, Generative Models, Variational Autoencoder, Semi Supervised Machine Learning.