EGFRNet: Transfer and Multi-task Learning Based on Graph Convolutional Network Toward Multi-target Drug Discovery Against Cancers for EGFR-Family Proteins

Saengnil, Natthakan (School: Kamnoetvidya Science Academy)
Virakarin, Puri (School: Kamnoetvidya Science Academy)

The epidermal growth factor receptor (EGFR) family proteins, consisting of HER1-4, are the most prominent causes of various human cancers, including breast, gastric, esophagus, lung, pancreatic, and bladder cancers. In addition, lung cancer treatments targeting HER1 protein can induce drug-resistant mutations. Currently, the standard drug discovery process encounters cost- and time-consuming challenges. However, the failures during research and development emerge from drug efficacy and safety problems. This scenario necessitates effective methods for discovering novel drugs to inhibit EGFR-family proteins. We herein developed models to predict bioactivity (pIC50) of kinase inhibitors against wild-type HER1-4 and mutant HER1 for multi-target drug discovery. Nevertheless, since the amount of training data is low, we employed transfer learning and multi-task learning techniques based on the LigEGFR model, the convolution spatial graph embedding network (C-SGEN) with deep neural network (DNN) algorithms trained on wild-type HER1. We performed the experiments by testing the average predictive performances through the RMSE and validated the models' robustness by using the y-scrambling technique. This study showed the transfer learning models and the multi-task learning models yielded higher predictive performance than traditional machine learning models. Our models provide a powerful strategy that may potentially help researchers to discover novel drugs against EGFR-family proteins. Moreover, these techniques can also be applied to virtual drug screening by using machine learning for multi-target drug development with other proteins through small-scale experimental data.

Awards Won:
First Award of $5,000