[BIOINFO 2025]

A Hybrid Convolution-Transformer Model for Drug-ATC Prediction Using Multiple Similarity Features

Gwang-Hyeon Yun[1], Jong-Hoon Park[1], Jae-Woo Chu[1],
Sang-Pil Cho[1], and Young-Rae Cho[1,2,*]

[1]*Division of Software, Yonsei University*
[2]*Division of Digital Healthcare, Yonsei University*
**Corresponding author: youngcho@yonsei.ac.kr*

Traditional de novo drug discovery faces significant hurdles, primarily characterized by high costs and a low probability of success. In response, drug repurposing has emerged as a promising and cost-effective strategy by identifying novel therapeutic applications for existing, approved drugs. Our work facilitated this process by developing a method for predicting a drug's Anatomical Therapeutic Chemical (ATC) code, thereby providing a systematic framework for inferring new therapeutic indications. To achieve this, we measured drug-drug similarities from multiple perspectives, including chemical structure, drug-drug interactions, and side effects. Furthermore, we proposed a novel similarity metric that leverages the hierarchical nature of ATC codes. These features are then used to construct a feature matrix representing drug-ATC associations, which is processed by a hybrid architecture combining a Convolutional layer and three Transformer encoders to learn complex interaction patterns. Our model achieved state-of-the-art performance as measured by AUROC and AUPRC across all four ATC levels, outperforming existing methods with a particularly strong improvement of 15.05% and 18.42% at level 4.