[BIOINFO 2024]

Transformer-Based Prediction of Drug Anatomical Therapeutic Chemical Code via Drug-Drug Interactions

Gwang-Hyeon Yun[1], Jong-Hoon Park[2], and Young-Rae Cho[1,3,*]

[1]*Division of Software, Yonsei University Mirae Campus*
[2]*Department of Computer Science, Yonsei University Mirae Campus*
[3]*Division of Digital Healthcare, Yonsei University Mirae Campus*
**Corresponding author: youngcho@yonsei.ac.kr*

The traditional drug development process takes 10 to 17 years and has a success rate of less than 10%, making it a high-risk industry. To mitigate these risks, there has been growing interest in drug repurposing through the use of artificial intelligence. Drug repurposing is an approach that involves re-evaluating existing drugs or those that failed in clinical trials to discover new therapeutic effects. Recently, research using the Anatomical Therapeutic Chemical (ATC) classification system has gained significant attention. The ATC code, developed by the World Health Organization (WHO), classifies drugs based on their target organ or system, pharmacological properties, and chemical structure, following a five-level hierarchical structure represented by up to seven alphanumeric characters. Predicting new ATC codes for drugs enables the discovery of novel therapeutic possibilities, contributing substantially to drug repurposing efforts.

In this study, we propose a novel framework called DDI-TR for predicting drug-ATC code associations. DDI-TR initially constructs a feature matrix by integrating drug similarity based on drug-drug interactions, ATC code similarities, and existing drug-ATC associations. This feature matrix is then transformed into a learnable format using convolutional operations, and a Transformer Encoder Layer is utilized to learn hidden relationships between drugs and ATC codes. Subsequently, the relationships between drugs and ATC codes are predicted using a Fully Connected Layer. We obtained drug-drug interaction and ATC code information from the Drugbank database, and through 10-fold cross-validation, we achieved maximum performance of AUROC 0.9802 and AUPRC 0.9786 for each layer of the ATC code, surpassing existing models.