**[BIOINFO 2024]**

# From the creation of new compounds using AI to the discovery of new drug candidate : Graph DF, Graph DTA

Choi, Jinkwang[1], Chankyu Park[2], JooHyun Kim[2], Rustamov Otabek[2] Cho, Toohyon[1,*]

*AIcloud., Ltd.*
**Corresponding author: ceo@aicloud.biz*

Recently, the use of artificial intelligence (AI) in drug development processes has been increasing among global pharmaceutical companies. AI offers the advantage of rapidly identifying potential candidates, but its predictive accuracy can decrease depending on the data used for training. In this study, we aimed to enhance prediction accuracy through hyperparameter tuning and cross-validation using multiple predictive models, in addition to leveraging the speed of individual models. Specifically, we generated 196,560 virtual compounds with guaranteed validity, uniqueness, and novelty using a diffusion model, which produces higher diversity than GANs and higher-quality data than VAEs. We then employed a GIN_Transformer model trained on BindingDatabase to predict 274 compounds with strong binding affinity to the TMPRSS2 target (IC50 ≤ 10 nM), using a voting method. After filtering for compounds with QED values above 0.5, 29 compounds were selected. After screening for synthetic feasibility, solubility, and ADMET prediction data to identify lead compounds, we plan to further identify candidate compounds using simulations. These candidates will be synthesized and subjected to in vitro experiments, followed by in vivo studies to identify compounds suitable for preclinical trials. For the generation of new compounds, approximately 35 hours were required when using an AMD EPYC 7543 32-Core Processor and an NVIDIA RTX A5000 GPU. Among these, 54,062 compounds had QED values higher than the ChEMBL database average of 0.404. In the target-compound binding prediction phase, the GIN_Transformer model can contribute to improving predictive accuracy by not only utilizing the structural information typical of graph models in DTI but also capturing sequence information patterns through the Transformer, offering a multidimensional perspective.