03-1

Novel QSAR approach for clearance prediction, combination DeepSnap-Deep Learning, and conventional machine learning

Hideaki Mamada[1, 2]
hideaki.mamada@jt.com

Yukihiro Nomura[1]
yukihiro.nomura@jt.com

Yoshihiro Uesawa[2]
uesawa@my-pharm.ac.jp

[1] Drug Metabolism and Pharmacokinetics Research Laboratories, Central Pharmaceutical Research Institute, Japan Tobacco Inc., 1-1, Murasaki-cho, Takatsuki, Osaka 569-1125, Japan
[2] Department of Medical Molecular Informatics, Meiji Pharmaceutical University, 2-522-1, Noshio, Kiyose-shi, Tokyo 204-858, Japan

Keywords: Clearance (CL), Quantitative structure-activity relationships (QSAR), DeepSnap-Deep Learning (DeepSnap-DL), ensemble model, consensus model

In drug discovery, there are some prediction targets for which the prediction accuracy by machine learning is not sufficient. Therefore, the development of new prediction models is required. In this study, rat clearance (CL) was selected as a challenging target because of poor prediction [1], and a new prediction model was developed. A classification model was constructed using 1545 in-house compounds for which rat CL data are available. The molecular descriptors calculated by Molecular Operating Environment (MOE), alvaDesc, and ADMET Predictor software were used to construct the prediction model. Molecular descriptors and random forest selected by DataRobot were used for conventional machine learning. The area under the curve (AUC) and accuracy (ACC) were 0.883 and 0.825, respectively. Conversely, compound images and Deep Learning were used for DeepSnap and Deep Learning (DeepSnap-DL) [2]. AUC and ACC were 0.905 and 0.832, respectively. The two models (conventional machine learning and DeepSnap-DL) were combined to develop a novel prediction model. The ensemble model using mean of the predicted probabilities from each model improved the evaluation scores (AUC=0.943 and ACC=0.874). Furthermore, using the results of the agreement between each classification as a consensus model resulted in higher ACC (=0.959). These combination models with a high level of predictive performance can be applied to rat CL as well as other pharmacokinetic parameters. These models will help the design of more rational compounds in drug discovery.

[1] McIntyre, T. A.; Han, C.; Davis, C. B. Prediction of animal clearance using naïve Bayesian classification and extended connectivity fingerprints, *Xenobiotica*, **2009,** *39*, 487–494. https://doi.org/10.1080/00498250902926906.

[2] Uesawa, Y. Quantitative structure–activity relationship analysis using deep learning based on a novel molecular image input technique, *Bioorganic & Medicinal Chemistry Letters,* **2018,** *28*, 3400–3403. https://doi.org/10.1016/j.bmcl.2018.08.032.