**Development of a Machine Learning-Based Predictive Model for Carrier Selection in Solid Dispersion Systems**

**Jin Woo Park**, Kwan Hyung Cho.

Department of Pharmacy, Inje University, Gimhae, Gyeongsangnam-do, Republic of Korea.

**Background and aims.** Solid dispersion (SD) technology is a widely used strategy to enhance the solubility and bioavailability of poorly water-soluble drugs. Polymer-based carriers play a key role in SD formulations. However, due to complex drug-excipient interactions, selecting the most suitable and effective carrier is time-intensive. This study aims to develop a predictive machine learning (ML) model that facilitates rational carrier selection for SD systems by analysing molecular structure-based compatibility data.

**Methods.** A comprehensive dataset was constructed using 246 published studies and 51 commercial SD products. Each active pharmaceutical ingredient (API) was paired with both soluble carriers and poorly soluble anti-carriers, resulting in 326 labelled API–carrier combinations. Polymeric materials were translated into SMILES notations, adapted for high molecular weight substances by defining repeat units with integer scaling. These structural representations were converted into molecular fingerprints suitable for machine learning input. Eight classification models (including knn, naive\_bayes, logit, svm, decision\_tree, random\_forest, mlp and xgb) were developed using 80 % of the dataset for training, and 20 % for validation. Model performance was evaluated based on accuracy, AUC, and MCC metrics.

**Results.** Logistic regression and multilayer perceptron (MLP) models showed high performance. Logistic regression and MLP achieved accuracy of 80.1 % and 81.0 %, with AUC and MCC values ​​slightly higher for the MLP model. Prediction evaluation showed moderate performance (60–75 %), especially limited in predicting incompatible pairs such as label 0. After training on label 0 data, the model accuracy was significantly improved to 85–90 %, confirming that machine learning performance is improved through continuous data training.

**Conclusion/Discussion.** Machine learning (ML) can effectively predict the solubility-based compatibility between API-Carrier used in SD. The developed model supports efficient SD formulation design and reduces empirical screening. Continuous expansion of training data will further improve the prediction performance and enable broader application in pharmaceutical formulation development.

**Acknowledgement.** This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (Grant number: NRF-2022R1A2C1003070).

**References:** After the discussion, you may include Acknowledgements section, for example, to thank funding sources or supporters of your work.

(1) Bao, Zeqing, et al. (2023) Advanced Drug Delivery Reviews 202:115108

(2) Bannigan, Pauric, et al. (2021) Advanced Drug Delivery Reviews 175:113806

(3) Murray, Jack D., et al. (2023) European Journal of Pharmaceutical Sciences 191:106562