Prediction Model of Clearance by a Novel Quantitative Structure–Activity Relationship Approach, Combination DeepSnap-Deep Learning and Conventional Machine Learning

Hideaki Mamada, Yukihiro Nomura, and Yoshihiro Uesawa*

Cite This: ACS Omega 2021, 6, 23570−23577

ABSTRACT: Some targets predicted by machine learning (ML) in drug discovery remain a challenge because of poor prediction. In this study, a new prediction model was developed and rat clearance (CL) was selected as a target because it is difficult to predict. A classification model was constructed using 1545 in-house compounds with rat CL data. The molecular descriptors calculated by Molecular Operating Environment (MOE), alvaDesc, and ADMET Predictor software were used to construct the prediction model. In conventional ML using 100 descriptors and random forest selected by DataRobot, the area under the curve (AUC) and accuracy (ACC) were 0.883 and 0.825, respectively. Conversely, the prediction model using DeepSnap and Deep Learning (DeepSnap-DL) with compound features as images had AUC and ACC of 0.905 and 0.832, respectively. We combined the two models (conventional ML and DeepSnap-DL) to develop a novel prediction model. Using the ensemble model with the mean of the predicted probabilities from each model improved the evaluation metrics (AUC = 0.943 and ACC = 0.874). In addition, a consensus model using the results of the agreement between classifications had an increased 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, pharmacological activity, and toxicity prediction. Therefore, these models will aid in the design of more rational compounds for the development of drugs.

INTRODUCTION

Quantitative structure–activity relationship (QSAR) analysis is a method to predict the absorption, distribution, metabolism, and excretion (ADME) parameters of small-molecule compounds based on their molecular structure. QSAR is used to predict ADME parameters including solubility, protein binding, permeability, blood-to-plasma concentration ratios, and metabolic stability, as well as in vivo pharmacokinetic (PK) parameters [clearance (CL), volume of distribution (Vd), and half-life]. The construction of QSAR models has mostly used molecular descriptors and fingerprints as features of compounds, as well as multiple regression, partial least-squares regression, random forest, support vector machines, neural networks, and XGBoost as algorithms. However, some prediction methods that use conventional ML for ADME parameters have a poor prediction accuracy. Recently, applying Deep Learning (DL) to the prediction of ADME parameters demonstrated accuracy improvements over other conventional ML methods. In these reports, molecular descriptors and fingerprints were used for DL as features of the compounds. However, the use of new features is expected to further improve the prediction accuracy in addition to DL. Uesawa recently reported a new method called DeepSnap, which uses images of compounds as features for DL. DeepSnap and Deep Learning (DeepSnap-DL) provided better predictions of toxicological targets including mitochondrial membrane potential disruption, constitutive androstane receptor (CAR), and aryl hydrocarbon receptor (AhR) compared with conventional ML. However, there have been no reports on ADME parameters using DeepSnap-DL. Therefore, constructing ADME parameters using DeepSnap-DL might have good prediction accuracy similar to that for toxicological targets.

Among ADME parameters, CL is an important PK parameter for drug discovery. CL in animal species, such as rats, is used to understand the relationship between compound exposure in animals and humans regarding PK, toxicity, and drug effects. It is desirable to obtain compounds that have an acceptable PK profile. However, most compounds do not have an acceptable PK profile at the early drug discovery stage. Grime et al. reported the efficient and cost-effective pursuit of candidate compounds.
with acceptable PK profiles.\textsuperscript{18} Pharmaceutical companies typically perform PK experiments in rats, including intravenous and oral administration, to determine whether compounds have an acceptable profile. To reduce the overall drug discovery cost, time, and animal usage, it is ideal to predict the rat PK profile before new chemical synthesis. QSAR can make predictions at early stages of drug development, even for virtual compounds, and can therefore help in the rational design of drug compounds.

Muegge \textit{et al}. and McIntyre \textit{et al}. reported QSAR models of rat clearance based on 6000–17,529 expanded in-house compounds.\textsuperscript{19,20} They constructed rat CL models using conventional ML. In their models, molecular descriptors and fingerprints were used as features of compounds. The naïve Bayesian method used by McIntyre \textit{et al}. and random forest or support vector machines (specific machine details not provided in Muegge \textit{et al}.) were used as algorithms. However, these prediction models did not show sufficient prediction performance [accuracy (ACC) of 0.74 and an area under the curve (AUC) of 0.82]. Although the prediction of rat CL is important in drug discovery, it is one of the evaluation targets that are difficult to predict by conventional ML. Therefore, in this study, we selected rat CL prediction as a difficult prediction target in drug discovery and developed a new prediction model using combination DeepSnap-DL and conventional ML to improve the evaluation metrics.

\section*{RESULTS}

\textbf{Separation of Compounds into Training and Test Datasets and Their Verification by Chemical Space Analysis.} Principal component analysis (PCA) was performed using a dataset of 1545 compounds with 11 representative molecular descriptors to confirm the correctness of the compound separation. It was previously reported that PCA could show the distribution of chemical space in the dataset.\textsuperscript{21} Components 1, 2, and 3 explained 62.3, 12.0, and 8.0\% of the variance, respectively. Figure 1 shows that the compounds were effectively separated into the training and test datasets.

\textbf{Construction of CL Prediction Models Using Molecular Descriptors by DataRobot.} The CL prediction models were constructed using 4795 molecular descriptors by DataRobot. First, random forest was selected as the algorithm based on the results of logloss of internal validation. Then, 100 molecular descriptors were selected from 4795 molecular descriptors using the permutation importance of random forest (Table S1). Second, these 100 descriptors were used to build a prediction model. Over 40 prediction models were constructed and evaluated. The top three models are shown in Table 1, and all results are shown in Table S2. Among these models, random forest showed the lowest logloss results. Based on this result, a final prediction model was constructed using 100\% training data by random forest. The results of the evaluation metrics for the test datasets are shown in Table 2. AUC, balanced accuracy (BAC), ACC, sensitivity, specificity, F-measure, precision, recall, and Matthews correlation coefficient (MCC) were 0.883, 0.817, 0.825, 0.772, 0.863, 0.784, 0.797, 0.772, and 0.638, respectively.

\textbf{Construction of CL Prediction Models by DeepSnap-DL.} The present study was conducted at four angles (65\textdegree{}, 85\textdegree{}, 105\textdegree{}, and 145\textdegree{}), with five learning rate conditions (0.0000001 – 0.001) and five maximum epoch conditions (15–300). In each condition, the epoch with the lowest loss [DeepSnap (Validation)] was selected as the epoch for calculating the evaluation metrics and the results of AUC [DeepSnap (Training)], AUC [DeepSnap (Validation)], and AUC (test) were calculated, respectively (Table S3). The AUC [DeepSnap (Validation)] was calculated for each condition, and the condition with the highest AUC [DeepSnap (Validation)] was selected as the final model in DeepSnap-DL. The highest AUC [DeepSnap (Validation)] in all conditions was observed at 145\textdegree{},
with a maximum epoch of 300 and a learning rate of 0.000001 (Table 3). At a learning rate of 0.000001 with a maximum epoch of 300 and a learning rate of 0.000001

Table 3. AUC of Validation Results by DeepSnap-DL

| learning rate | max. epoch | 145° | 105° | 85° | 65° |
|---------------|------------|------|------|-----|-----|
| 0.001         | 15         | 0.812 | 0.779 | 0.799 | 0.500 |
|               | 30         | 0.771 | 0.787 | 0.798 | 0.500 |
|               | 60         | 0.500 | 0.789 | 0.500 | 0.820 |
|               | 100        | 0.500 | 0.791 | 0.500 | 0.788 |
|               | 300        | 0.818 | 0.777 | 0.500 | 0.782 |
| 0.0001        | 15         | 0.850 | 0.793 | 0.809 | 0.790 |
|               | 30         | 0.831 | 0.803 | 0.797 | 0.770 |
|               | 60         | 0.801 | 0.840 | 0.817 | 0.791 |
|               | 100        | 0.801 | 0.813 | 0.806 | 0.791 |
|               | 300        | 0.814 | 0.820 | 0.842 | 0.802 |
| 0.00001       | 15         | 0.879 | 0.882 | 0.873 | 0.831 |
|               | 30         | 0.887 | 0.868 | 0.833 | 0.837 |
|               | 60         | 0.879 | 0.824 | 0.837 | 0.860 |
|               | 100        | 0.868 | 0.841 | 0.839 | 0.855 |
|               | 300        | 0.867 | 0.848 | 0.837 | 0.857 |
| 0.000001      | 15         | 0.740 | 0.851 | 0.864 | 0.875 |
|               | 30         | 0.847 | 0.870 | 0.883 | 0.844 |
|               | 60         | 0.867 | 0.884 | 0.893 | 0.887 |
|               | 100        | 0.877 | 0.889 | 0.893 | 0.879 |
|               | 300        | 0.897 | 0.897 | 0.891 | 0.863 |
| 0.0000001     | 15         | 0.644 | 0.625 | 0.670 | 0.701 |
|               | 30         | 0.650 | 0.661 | 0.728 | 0.820 |
|               | 60         | 0.666 | 0.729 | 0.828 | 0.856 |
|               | 100        | 0.689 | 0.826 | 0.850 | 0.871 |
|               | 300        | 0.844 | 0.869 | 0.883 | 0.844 |

“MD-based method, molecular descriptor-based method; DeepSnap-DL results are shown in Table 4a,b. Based on these results, a consensus model was constructed using the results of the agreement (Table 4c). Evaluation metrics showed that BAC, ACC, sensitivity, specificity, F-measure, precision, recall, and MCC were 0.958, 0.959, 0.953, 0.963, 0.948, 0.943, 0.953, and 0.915, respectively (Table 2). The number of predictable compounds decreased from 309 to 214. However, these results showed that the consensus model was highly accurate for all the evaluation metrics.

### DISCUSSION

During drug discovery, prediction models are constructed for various targets such as toxicity and ADME parameters, but the prediction performance of these models is insufficient for some targets. Therefore, a new prediction model that has a high level of predictive performance is desired. In this study, we focused on the prediction of rat CL as an important and difficult prediction target in drug discovery. To improve the prediction performance, we developed a new prediction model with DeepSnap-DL, which uses images for ML.

For rat CL dataset creation, compounds were separated into training and test sets (Table 5 and Figure 1). To ensure unbiased
segregation, the PCA analysis was conducted using 11 representative molecular descriptors (Table S4), which are generally considered important for synthetic expansion.21−24 We also examined the distribution of each training set and test set for the 11 descriptors (Figure S1). As shown in Figure 1 and Figure S1, the separation was well balanced and the cumulative contribution ratio of PCA from 1 to 3 was 82.31%.

For rat CL prediction, two models have been reported to date, although a direct comparison is difficult because both are different in-house compounds.19,20 However, the prediction performance of both models was low, with an ACC of 0.74 and an AUC of 0.82.19,20 These models also used molecular descriptors and fingerprints as features of compounds, as well as random forest (or support vector machines) and naive Bayesian as algorithms. In this study, we used molecular descriptors obtained from three software, Molecular Operating Environment (MOE), alvaDesc, and ADMET Predictor, and constructed a model using DataRobot, which allows multiple algorithms to be considered simultaneously as conventional ML. As a result, evaluation metrics calculated an ACC of 0.825 and an AUC of 0.883 (Table 2). Although it is difficult to make a direct comparison because of the different conditions used, we constructed a prediction model that surpassed previous models by adopting multiple software for molecular descriptors and multiple algorithms.

We developed a prediction model using DeepSnap, which uses images as features of compounds and DL as an algorithm. First, we examined the hyperparameters of DeepSnap-DL for predicting rat CL. The results for all hyperparameter combination conditions in this study are shown in Table S3, and the AUC results for internal validation are shown in Table 3. The condition of 145°, learning rate of 0.000001, and maximum epoch of 300 showed the highest value of AUC [DeepSnap (Validation)] (Table 3). We evaluated the prediction performance of the test sets using this condition for the final model of DeepSnap-DL. As shown in Table 2, the ACC was 0.832 and the AUC was 0.905, which were higher than when using the molecular descriptor-based method (conventional ML). It was reported that DeepSnap-DL had a higher prediction performance than conventional ML for multiple toxicity targets of progesterone receptor, CAR, and AhR.15−17 Although DeepSnap-DL has only been used for toxicity targets, it also had high prediction performance for PK parameters.

In this study, we focused on the multiple QSAR model to improve the prediction performance further. The multiple QSAR models [ensemble learning, combinatorial (combi) QSAR, and consensus classification] were used to improve prediction and obtain stable results by combining different features or algorithms.22−24 Various multiple QSAR models have been reported to date. Combi QSAR is a method for constructing models by combining molecular descriptors in multiple commercial software and multiple algorithms (k-nearest neighbor, support vector machine, decision trees, and random forest).25−28 It was reported that high prediction accuracy and stable results were obtained by using these methods.25−28 Furthermore, Brownfield et al. proposed a prediction method that combined three class systems by a fusion process as a consensus classification.29 Kim et al. and Wang et al. showed an improvement in prediction accuracy by using a molecular descriptor and the parameter of transporter as a biological descriptor for the prediction model.30,31 From these findings, it was expected that the use of multiple prediction models and different types of features might improve the prediction performance. Therefore, we investigated the combination of prediction models using a molecular descriptor-based model and DeepSnap-DL. We developed two multiple QSAR models, the ensemble model and consensus model. For the ensemble model, the average of the prediction probabilities obtained from the molecular descriptor-based model and DeepSnap-DL was used. As a result, the AUC and ACC were improved because these scores increased from 0.883−0.905 to 0.943 and from 0.825−0.832 to 0.874, respectively (Table 2). To confirm that this was not a coincidence, we examined different test partitions (Figure S2), which showed that the average of the prediction probabilities improved the prediction performance (Table S5). For the consensus model, only the results that agreed with the molecular descriptor-based method and DeepSnap-DL were used. The evaluated number of compounds decreased from 309 to 214 because different prediction results could not be used. However, the accuracy of prediction using the consensus model was improved from 0.825−0.832 to 0.959 (Table 2). We examined different test partitions as was done for the ensemble model and found that using the consensus model improved the prediction performance (Figure S2 and Table S5). Although the consensus model has been shown to have the highest prediction accuracy, it is not possible to evaluate all test compounds. In fact, the number of compounds that can be evaluated by the consensus model has been reduced to 214−226 (Table 2 and Table S5). In the early stages of drug screening, a large number of compounds need to be evaluated comprehensively without omission. The model that can evaluate all compounds is suitable, so further efforts are needed for practical application. In this study, the ensemble model and the consensus model improved the prediction performance. To the best of our knowledge, this is the first report showing an improvement in prediction performance using images and molecular descriptors of compounds. The reason for the improved prediction performance using this combination is that the recognition of compounds in the image space and molecular descriptor space is different. This suggests that the high prediction model was achieved by using information in each space.

■ CONCLUSIONS

In this study, we constructed a novel combination model using different types of compound features that had high performance for rat CL prediction. Although this combination model was effective for rat CL, it may be applicable to other pharmacokinetic parameters, toxicity, and pharmacological activity. This combination QSAR method enables virtual screening from a library of compounds and accelerates drug discovery. Furthermore, this model is expected to enable the construction of a prediction model that outperforms previous models for drug discovery as well as other compound-based targets. Therefore, it is expected to be widely applicable in various fields when prediction performance is an issue.

■ EXPERIMENTAL SECTION

Experimental Data. All procedures for the animal experiments were approved by the Animal Ethics Committee of Japan Tobacco Inc., Central Pharmaceutical Research Institute. The CL of compounds that were synthesized in multiple in-house projects and subjected to rat PK was obtained from the in-house database. All results were obtained after the intravenous administration of compounds to rats at doses of 0.03−10.0
mg/kg. CL was estimated after the intravenous administration by the noncompartmental analysis of individual plasma concentration–time profiles. In this study, the results of the CL of 1545 in-house compounds were used to construct the prediction models. The threshold for the classification of CL values was 1 L/h/kg, which is approximately 30% of the hepatic blood flow rate in rats. Subsequently, the data of compounds for water molecules and counter ions were eliminated by the processing of disposal salts. 

The threshold was set at 1 L/h/kg, which is equivalent to 70% BA. McIntyre et al. developed similar prediction models, but they used 70% of hepatic blood flow as their threshold. This is equivalent to 30% BA when FaFg = 1. However, it is difficult to obtain a compound with FaFg = 1 early in drug discovery, and the BA is easily likely to be below 30%. Therefore, in this study, the threshold was set at 1 L/h/kg, which is equivalent to 70% BA.

**Calculation of Molecular Descriptors.** The structural data of compounds for water molecules and counter ions were eliminated by the processing of disposal salts. Subsequently, the 3D structure of each compound was optimized using “Rebuild 3D” and the force field calculations (amber-10: EHT) were conducted in MOE version 2019.0102 (MOLIS Inc., Tokyo, Japan). Structural descriptors were calculated employing MOE, alvaDesc (1.0.16) (Alvascience srl, Lecco, Italy), and ADMET Predictor (9.5.0.16) (Simulations Plus, New York, NY, USA). At the time of descriptor generation, descriptors of string type were removed in ADMET Predictor and descriptors of variance 0 were removed in alvaDesc. Overall, 4795 descriptors were selected for further analysis.

**Separation of Compounds into Training and Test Sets and Their Verification by Chemical Space Analysis.** After applying stratified random sampling, the compounds in the dataset were separated randomly into a training set and test set at a ratio of 4:1 (Table 5). To investigate the chemical space, 11 molecular parameters were used as reported previously with JMP Pro software 14.3.0 (SAS Institute Inc., Cary, NC, USA) PCA.21 The parameters included molecular weight, SlogP (log octanol/water partition coefficient), topological polar surface area (TPSA), h_logD [log octanol/water distribution coefficient (pH = 7)], h_pKa [acidity (pH = 7)], h_pKb [basicity (pH = 7)], a_acc (number of H-bond acceptor atoms), a_don (number of H-bond donor atoms), a_aro (number of aromatic atoms), b_ar (number of aromatic bonds), and b_rotN (number of rotatable bonds). The principal components were calculated from 1 to 3.

**Construction of Rat CL Models Based on Molecular Descriptors.** Model construction based on 4795 molecular descriptors was performed using DataRobot (SaaS, DataRobot, Tokyo, Japan). All analyses were conducted from 29 July 2020 to 31 July 2020. DataRobot automatically performs a modeling competition in which a wide selection of algorithm and data preprocessing techniques compete with one another as reported previously.35,36 Prior to training, 20% of the training dataset was randomly selected as the holdout and excluded from training. Five-fold cross-validation was implemented and the partitions were determined with stratified sampling. After the selection of models based on logloss scores of internal validation, molecular descriptors were selected from 4795 molecular descriptors to 100 using the permutation importance. Using these 100 selected molecular descriptors, over 40 models were created and random forest was selected based on validation results as the final algorithm. Following logloss scores of internal validation, the best model was constructed using 100% of the training data. This final model was used to calculate the prediction accuracy of the test sets (Figure 2).

**DeepSnap.** 3D structures were saved in an SDF file format after "Rebuild 3D" (refer to Calculation of Molecular Descriptors). The 3D chemical structures were depicted as 3D ball-and-stick models with different colors corresponding to different atoms by Jmol, open-source Java viewer software for the 3D molecular modeling of chemical structures, as previously reported.11–17 The 3D chemical structures were captured automatically as snapshots with user-defined angle increments with respect to the x, y, and z axes. In this study, a four-angle increment was used: (65°, 65°, 65°), (85°, 85°, 85°), (105°, 105°, 105°), and (145°, 145°, 145°). Other parameters for the DeepSnap depiction process were set based on previous studies12–17 as follows: image pixel size, 256 × 256; molecule number per SDF file to split into, 100; zoom factor (%), 100; atom size for van der Waals radius (%), 23; bond radius (mÅ), 15; minimum bond distance, 0.4; bond tolerance, 0.8. The snapshots were saved as 256 × 256 pixel-resolution PNG files (RGB). These were divided into three types of datasets [DeepSnap (Training), DeepSnap (Validation), and test] as described below. After sorting based on CL, training datasets were separated randomly into DeepSnap (Training) or DeepSnap (Validation) at a ratio of 3:1.

**Deep Learning.** All the two-dimensional (2D) PNG images produced by DeepSnap were resized by utilizing NVIDIA DL GPU Training System (DIGITS) version 6.0.0 software (NVIDIA, Santa Clara, CA, USA) on four-GPU systems, Te2D sla-V100 (32 GB), with a resolution of 256 × 256 pixels as input data, as previously reported.11–17 To rapidly train and fine-tune the highly accurate Convolutional Neural Network (CNN) using the input DeepSnap (Training) and DeepSnap (Validation) datasets based on the image classification and by building the pretrained prediction model, we used a pretrained...
open-source DL model, Caffe,\textsuperscript{35} and open-source software on the Ubuntu distribution 16.04LTS. In this study, the deep CNN architecture was GoogLeNet and Adam was used for optimization. In the DeepSnap-DL method, the prediction models were constructed by DeepSnap (Training) datasets using 15–300 epochs with 1 snapshot interval in each epoch, 1 validation interval in each epoch, 1 random seed, a learning rate of 0.0000001–0.001, and a batch size of default in DIGITS. Among the epochs, the lowest loss value in the DeepSnap (Validation) datasets, which is the error rate between the results obtained from the DeepSnap (Validation) datasets and the corresponding labeled dataset, was selected for the subsequent examination of prediction using the test set. The probability of the prediction results with the lowest minimum loss [DeepSnap (Validation)] value was analyzed. The probabilities for each image of one molecule captured at different angles with respect to the x, y, and z axes using DeepSnap-DL were calculated. The medians of each of these predicted values were used as the representative values for target molecules as previously reported.\textsuperscript{11–17}

**Combination DeepSnap-DL and Conventional ML.** In this study, we investigated the combination of DeepSnap-DL and conventional ML using two methods. The first method was the average of the prediction probabilities. The prediction probabilities obtained by DeepSnap-DL and conventional ML were averaged, and this average value was used as the prediction probability of the new prediction model (ensemble model) (Figure 2). The second method used a prediction model that adopted the results of the agreement between DeepSnap-DL and conventional ML (consensus model) (Figure 2).

**Evaluation of the Predictive Model.** The performance of each model in predicting rat CL was evaluated in terms of the following metrics: AUC, BAC, ACC, sensitivity, specificity, F-measure, precision, recall, and MCC calculated using KNIME (4.1.4) (KNIME, Konstanz, Germany). These performance metrics were defined as follows:

\[ \text{BAC} = \frac{\text{sensitivity} + \text{specificity}}{2} \]

\[ \text{ACC} = \frac{(TP + TN)/(TP + FP + TN + FN)}{\sum \text{TPs} / (\sum \text{TPs} + \sum \text{FNs})} \]

\[ \text{sensitivity} = \frac{\sum \text{TPs}}{\sum \text{TPs} + \sum \text{FNs}} \]

\[ \text{specificity} = \frac{\sum \text{TPs}}{\sum \text{TPs} + \sum \text{FNs}} \]

\[ \text{F-measure} = \frac{2(\text{recall})(\text{precision})}{(\text{recall} + \text{precision})} \]

\[ \text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \]

\[ \text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \]

\[ \text{MCC} = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \]

where TP, FN, TN, and FP denote true positive, false negative, true negative, and false positive, respectively. To determine the optimal cutoff point for the definition of TP, FN, TN, and FP, the method of maximizing sensitivity (1 – specificity), termed the Youden index,\textsuperscript{36,37} was used. The index has a value ranging from 0 to 1, where 1 represents over 1 L/h/kg and 0 represents less than 1 L/h/kg for rat CL.

### ASSOCIATED CONTENT

**Supporting Information**

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.1c03689.

- Physicochemical property distribution of compounds (n = 1545) (Figure S1); partition patterns for the model (Figure S2) (PDF)
- One hundred molecular descriptors selected by random forest (Table S1); internal validation results using molecular descriptor-based methods (Table S2); prediction performances with 145° on DeepSnap-Deep Learning (Table S3a); prediction performances with 105° on DeepSnap-Deep Learning (Table S3b); prediction performances with 85° on DeepSnap-Deep Learning (Table S3c); prediction performances with 65° on DeepSnap-Deep Learning (Table S3d); details of 11 representative descriptors (Table S4); prediction performances at different partition pattern 1 (Table S5a); prediction performances at different partition pattern 2 (Table S5b); prediction performances at different partition pattern 3 (Table S5c); prediction performances at different partition pattern 4 (Table S5d) (XLSX)

### AUTHOR INFORMATION

**Corresponding Author**

Yoshihiro Uesawa — Department of Medical Molecular Informatics, Meiji Pharmaceutical University, Tokyo 204-858, Japan; orcid.org/0000-0002-5773-991X; Phone: +81-42-495-8983; Email: uesawa@my-pharm.ac.jp; Fax: +81-42-495-8983

**Authors**

Hideaki Mamada — Department of Medical Molecular Informatics, Meiji Pharmaceutical University, Tokyo 204-858, Japan; Drug Metabolism and Pharmacokinetics Research Laboratories, Central Pharmaceutical Research Institute, Japan Tobacco Inc., Osaka 569-1125, Japan; orcid.org/0000-0002-5433-7042

Yukihiro Nomura — Drug Metabolism and Pharmacokinetics Research Laboratories, Central Pharmaceutical Research Institute, Japan Tobacco Inc., Osaka 569-1125, Japan

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.1c03689

**Funding**

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

**Notes**

The authors declare no competing financial interest.

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript. All authors contributed to the study conception and design. Data collection was performed by H.M. Y.U., H.M., and Y.N. developed the QSAR models with experimental data. The first draft of the manuscript was written by H.M. and all authors commented on previous versions of the manuscript.

**ACKNOWLEDGMENTS**

We gratefully acknowledge the work of past and present members of our laboratory.


**ABBREVIATIONS**

QSAR, quantitative structure–activity relationship; ADME, absorption, distribution, metabolism, and excretion; PK, pharmacokinetics; CL, clearance; Vd, volume of distribution; ML, machine learning; DL, Deep Learning; DeepSnap-DL, DeepSnap and Deep Learning; CAR, constitutive androstane receptor; AhR, aryl hydrocarbon receptor; ACC, accuracy; AUC, area under the curve; MCC, Matthews correlation coefficient; AVG Blender, average Blender; ENET blender, Elastic-Net Blender; MD-based method, molecular descriptor-based method; MOE, Molecular Operating Environment; combi, combinatorial; BA, bioavailability; Fa, fraction absorbed; Fg, fraction intestinal availability; SlogP, log octanol/water partition coefficient; TPSA, topological polar surface area; h_logD, octanol/water distribution coefficient; a_acc, number of H-bond acceptor atoms; a_don, number of H-bond donor atoms; a_ar, number of aromatic bonds; b_rotN, number of rotatable bonds; 2D, two-dimensional; DIGITS, NVIDIA DL GPU Training System; CNN, Convolutional Neural Network

**REFERENCES**

(1) Wanchana, S.; Yamashita, F.; Hashida, M. Quantitative Structure/Property Relationship Analysis on Aqueous Solubility Using Genetic Algorithm-Combined Partial Least Squares Method. *Pharmazie 2002*, *57*, 127–129.

(2) Watanabe, R.; Esaki, T.; Kawashima, H.; Natsume-Kitatani, Y.; Nago, C.; Ohashi, R.; Mizuguchi, K. Predicting Fraction Unbound in Human Plasma from Chemical Structure: Improved Accuracy in the Low Value Ranges. *Mol. Pharmaceutics* 2018, *15*, 5302–5311. (accessed date; 14, April, 2021)

(3) Fujiwara, S.-i.; Yamashita, F.; Hashida, M. Prediction of Caco-2 Cell Permeability Using a Combination of MO-Calculation and Neural Network. *Int. J. Pharm.* 2002, *237*, 95–105. (accessed date; 14, April, 2021)

(4) Mamada, H.; Iwamoto, K.; Nomura, Y.; Uesawa, Y. Predicting Blood-to-Plasma Concentration Ratios of Drugs from Chemical Structures and Volumes of Distribution in Humans. *Mol. Diversity* 2021, *25*, 1261–1270. (accessed date; 16, August, 2021)

(5) Shen, M.; Xiao, Y.; Golbraikh, A.; Gombar, V. K.; Tropsha, A. Development and Validation of k-Nearest-Neighbour QSRR Models of Metabolic Stability of Drug Candidates. *J. Med. Chem.* 2003, *46*, 3013–3020. (accessed date; 14, April, 2021)

(6) Wang, Y.; Liu, H.; Fan, Y.; Chen, X.; Yang, Y.; Zhu, L.; Zhao, J.; Chen, Y.; Zhang, Y. In Silico Prediction of Human Intravenous Pharmacokinetic Parameters with Improved Accuracy. *J. Chem. Inf. Model.* 2019, *59*, 3968–3980. (accessed date; 14, April, 2021)

(7) Lombardo, F.; Obach, R. S.; Varma, M. V.; Stringer, R.; Berellini, G. Clearance Mechanism Assignment and Total Clearance Prediction in Human Based upon in Silico Models. *J. Med. Chem.* 2014, *57*, 4397–4405. (accessed date; 14, April, 2021)

(8) Ma, J.; Sheridan, R. P.; Liaw, A.; Dahl, G. E.; Svetnik, V. Deep Neural Nets as a Method for Quantitative Structure-Activity Relationships. *J. Chem. Inf. Model.* 2015, *55*, 263–274. (accessed date; 14, April, 2021)

(9) Korotcov, A.; Tkachenko, V.; Russo, D. P.; Ekins, S. Comparison of Deep Learning with Multiple Machine Learning Methods and Metrics Using Diverse Drug Discovery Data Sets. *Mol. Pharmaceutics* 2017, *14*, 4462–4475. (accessed date; 14, April, 2021)

(10) Koutrouskas, A.; Monaghan, K. J.; Li, X.; Huan, J. Deep-Learning: Investigating Deep Neural Networks Hyper-Parameters and Comparison of Performance to Shallow Methods for Modeling Bioactivity Data. *Aust. J. Chem.* 2017, *9*, 42. (accessed date; 14, April, 2021)

(11) Uesawa, Y. Quantitative Structure–Activity Relationship Analysis Using Deep Learning Based on a Novel Molecular Image Input Technique. *Bioorg. Med. Chem. Lett.* 2018, *28*, 3400–3403. (accessed date; 14, April, 2021)

(12) Matsuzaka, Y.; Uesawa, Y. A Molecular Image-Based Novel Quantitative Structure-Activity Relationship Approach, Deepsnap-Deep Learning and Machine Learning. *Curr. Issues Mol. Biol.* 2022, *42*, 455–472. (accessed date; 1, July, 2021)

(13) Matsuzaka, Y.; Uesawa, Y. Optimization of a Deep-Learning Method Based on the Classification of Images Generated by Parameterized Deep Snap a Novel Molecular-Image-Input Technique for Quantitative Structure-Activity Relationship (QSAR) Analysis. *Front. Bioeng. Biotechnol.* 2019, *7*, 65. (accessed date; 14, April, 2021)

(14) Matsuzaka, Y.; Uesawa, Y. Molecular Image-Based Prediction Models of Nuclear Receptor Agonists and Antagonists Using the DeepSnap-Deep Learning Approach with the Tox21 10K Compound Library. *Molecules* 2020, *25*, 2764. (accessed date; 14, April, 2021)

(15) Matsuzaka, Y.; Uesawa, Y. DeepSnap-Deep Learning Approach Predicts Progesterone Receptor Antagonist Activity With High Performance. *Front. Bioeng. Biotechnol.* 2020, *7*, 485. (accessed date; 14, April, 2021)

(16) Matsuzaka, Y.; Uesawa, Y. Prediction Model with High-Performance Constitutive Androstane Receptor (CAR) Using DeepSnap-Deep Learning Approach from the Tox21 10K Compound Library. *Int. J. Mol. Sci.* 2019, *20*, 4855. (accessed date; 14, April, 2021)

(17) Matsuzaka, Y.; Hosaka, T.; Ogaito, A.; Yoshinari, K.; Uesawa, Y. Prediction Model of Aryl Hydrocarbon Receptor Activation by a Novel QSAR Approach, DeepSnap-Deep Learning. *Molecules* 2020, *25*, 1317. (accessed date; 14, April, 2021)

(18) Grime, K. H.; Barton, P.; McGinnity, D. F. Application of in Silico, in Vitro and Preclinical Pharmacokinetic Data for the Effective and Efficient Prediction of Human Pharmacokinetics. *Mol. Pharmaceutics* 2013, *10*, 1191–1206. (accessed date; 14, April, 2021)

(19) Muegge, I.; Bergner, A.; Kriegl, J. M. Computer-Aided Drug Design at Boehringer Ingelheim. *J. Comput.-Aided Mol. Des.* 2017, *31*, 275–285. (accessed date; 14, April, 2021)

(20) McIntyre, T. A.; Han, C.; Davis, C. B. Prediction of Animal Clearance Using Naive Bayesian Classification and Extended Connectivity Fingerprints. *Xenobiota* 2009, *39*, 487–494. (accessed date; 14, April, 2021)

(21) Ohashi, R.; Watanabe, R.; Esaki, T.; Taniguchi, T.; Torimoto-Katori, N.; Watanabe, T.; Ogasawara, Y.; Takahashi, T.; Tsukimoto, M.; Mizuguchi, K. Development of Simplified in Vitro P-Glycoprotein Substrate Assay and in Silico Prediction Models to Evaluate Transport Potential of P-Glycoprotein. *Mol. Pharmaceutics* 2019, *16*, 1851–1863. (accessed date; 14, April, 2021)

(22) Zhang, L.; Tan, J.; Han, D.; Zhu, H. From Machine Learning to Deep Learning: Progress in Machine Intelligence for Rational Drug Discovery. *Drug Discovery Today* 2017, *22*, 1680–1685. (accessed date; 14, April, 2021)

(23) Rokach, L. Ensemble-Based Classifiers. *Artif. Intell. Rev.* 2010, *33*, 1–39. (accessed date; 14, April, 2021)

(24) Kottler, J.; Hatef, M.; Duin, R. P. W.; Matas, J. On Combining Classifiers. *IEEE Trans. Pattern Anal. Mach. Intell.* 1998, *20*, 226–239. (accessed date; 14, April, 2021)

(25) Kovatcheva, A.; Golbraikh, A.; Oloff, S.; Xiao, Y.-D.; Zheng, W.; Wolschmann, P.; Buchbauer, G.; Tropsha, A. Combinatorial QSAR of Ambergris Fragrance Compounds. *J. Chem. Inf. Comput. Sci.* 2004, *44*, 582–595. (accessed date; 14, April, 2021)

(26) de Cerqueira Lima, P.; Golbraikh, A.; Oloff, S.; Xiao, Y.; Tropsha, A. Combinatorial QSAR Modeling of P-Glycoprotein Substrates. *J. Chem. Inf. Model.* 2006, *46*, 1245–1254. (accessed date; 14, April, 2021)

(27) Zhang, L.; Zhu, H.; Oprea, T. I.; Golbraikh, A.; Tropsha, A. QSAR Modeling of the Blood-Brain Barrier Permeability for Diverse Organic Compounds. *Pharm. Res.* 2008, *25*, 1902–1914. (accessed date; 14, April, 2021)

(28) Sprague, B.; Shi, Q.; Kim, M. T.; Zhang, L.; Sedykh, A.; Ichiiishi, E.; Tokuda, H.; Lee, K.-H.; Zhu, H. Design, Synthesis and Experimental Validation of Novel Potential Chemopreventive Agents Using Random
Forest and Support Vector Machine Binary Classifiers. *J. Comput.-Aided Mol. Des.* 2014, 28, 631–646. (accessed date; 14, April, 2021)

(29) Brownfield, B.; Lemos, T.; Kalivas, J. H. Consensus Classification Using Non-Optimized Classifiers. *Anal. Chem.* 2018, 90, 4429–4437. (accessed date; 14, April, 2021)

(30) Kim, M. T.; Sedykh, A.; Chakravarti, S. K.; Saiakhov, R. D.; Zhu, H. Critical Evaluation of Human Oral Bioavailability for Pharmaceutical Drugs by Using Various Cheminformatics Approaches. *Pharm. Res.* 2014, 31, 1002–1014. (accessed date; 14, April, 2021)

(31) Wang, W.; Kim, M. T.; Sedykh, A.; Zhu, H. Developing Enhanced Blood-Brain Barrier Permeability Models: Integrating External Bio-Assay Data in QSAR Modeling. *Pharm. Res.* 2015, 32, 3055–3065. (accessed date; 14, April, 2021)

(32) Davies, B.; Morris, T. Physiological Parameters in Laboratory Animals and Humans. *Pharm. Res.* 1993, 10, 1093–1095. (accessed date; 14, April, 2021)

(33) Tsuzuki, S.; Fujitsuka, N.; Horiiuchi, K.; Ijichi, S.; Gu, Y.; Fujitomo, Y.; Takahashi, R.; Ohmagari, N. Factors Associated with Sufficient Knowledge of Antibiotics and Antimicrobial Resistance in the Japanese General Population. *Sci. Rep.* 2020, 10, 3502. (accessed date; 14, April, 2021)

(34) Muhlestein, W. E.; Akagi, D. S.; Davies, J. M.; Chambless, L. B. Predicting Inpatient Length of Stay after Brain Tumor Surgery: Developing Machine Learning Ensembles to Improve Predictive Performance. *Neurosurgery* 2019, 85, 384–393. (accessed date; 14, April, 2021)

(35) Jia, Y.; Shelhamer, E.; Donahue, J.; Karayev, S.; Long, J.; Girshick, R.; Guadarrama, S.; Darrell, T. Caffe: Convolutional Architecture for Fast Feature Embedding. In *Proceedings of the 22nd ACM international conference on Multimedia*; 2014, 675–678, DOI: 10.1145/2647868.2654889 (accessed date; 14, April, 2021).

(36) Liang, K.; Wang, C.; Yan, F.; Wang, L.; He, T.; Zhang, X.; Li, C.; Yang, W.; Ma, Z.; Ma, A.; Hou, X.; Chen, L. HbA1c Cutoff Point of 5.9% Better Identifies High Risk of Progression to Diabetes among Chinese Adults: Results from a Retrospective Cohort Study. *J. Diabetes Res.* 2018, 2018, 7486493. (accessed date; 14, April, 2021)

(37) Yun, J.-H.; Chun, S.-M.; Kim, J.-C.; Shin, H.-I. Obesity Cutoff Values in Korean Men with Motor Complete Spinal Cord Injury: Body Mass Index and Waist Circumference. *Spinal Cord* 2019, 57, 110–116. (accessed date; 14, April, 2021)