

機械学習を用いた薬物動態パラメータ予測モデル構築に関する研究

Construction of Models for Predicting Pharmacokinetic Parameters Using
Machine Learning

令和4年度 論文博士申請者 俣田 秀章 (Mamada, Hideaki)

薬物動態の理論に基づいた血中薬物濃度予測は、創薬における化合物探索および臨床における適切な薬物治療のために重要である。薬物の効果発現には十分な血中濃度が必要であるが、血中濃度が高すぎると副作用が発現することから、血中濃度の正確な予測がもとめられる。薬物動態研究において、クリアランス(CL)、分布容積(Vd)など複数の薬物動態パラメータが血液中薬物濃度に基づいて算出される。これらの薬物動態パラメータは、創薬過程において動物実験もしくは *in vitro* 試験に基づいて評価され、最終的にヒトを対象とした臨床試験を経て確定される。

近年、経済的・時間的コストの削減、動物実験の削減、および予測精度向上のため、創薬過程における薬物動態パラメータの予測に定量的構造活性相関(QSAR)解析が利用されている。QSAR 解析は化合物の構造をコンピュータによって解析可能な特徴量に変換して数理モデルを構築する手法である。本手法では訓練セットを用いて予測モデルを構築し、テストセットを用いて予測モデルの精度を評価する。これまでに、従来では実験的に獲得されてきた溶解度、代謝安定性、タンパク結合率など、多くの薬物動態パラメータが QSAR 解析によって予測モデル化され、報告されている。しかし、QSAR 解析の報告が極端に少ない、もしくは予測精度が不十分で

あるために実用的な予測が困難な薬物動態パラメータも存在する。そこで、本研究では臨床上重要なパラメータであるにも関わらず予測モデルの報告がほとんどない、血液中-血漿中薬物濃度比 (Rb) の予測モデルを構築した。さらに、予測精度が不十分な薬物動態パラメータとして創薬段階で重要な CL について、予測精度向上のための新たな手法の開発を試みた。

1. 化学構造および分布容積を用いたヒト血液中-血漿中薬物濃度比の予測モデルの構築¹⁾

Rb は臨床において薬剤の投与方針を決定する上で重要なパラメータの一つである。しかしながら、Rb 予測に QSAR 解析を適用した研究は Paixão らの報告のみである²⁾。そこで、QSAR 解析における化学構造情報と薬物動態パラメータを併用することによって、より高精度なヒト Rb 予測モデルの構築を検討した。Paixão らは Rb が報告されている 93 化合物での解析を行ったが、本研究では複数の文献値に加えて 62 化合物の Rb を実測することにより、289 化合物(訓練セット：193 化合物、テストセット：96 化合物に層化抽出法で無作為に分割)に対する Rb を取得し解析に用いた。はじめに、専用ソフトウェアを用いて分子記述子を算出し、それぞれの記述子の共線性および分散により 1777 種の分子記述子を選択した。これらの分子記述子からアルゴリズムの一つである Random Forest によって予測モデルを構築したところ、相関係数(r)の平均値および標準偏差は各々0.477 および 0.064 であった。次に、予測精度改善を目指し、説明変数として分子記述子に加えて実験値の使用を試みた。予測モデルで使用する実験値として、ヒトの薬物動態パラメータである Vd、CL、平均滞留時間、タンパク結合率を文献および *in vitro* 実験により取得した。Rb と薬物動態パラメータの相関性を検討した結果、Vd に高度な相関を認めた (r=0.47)。そこ

で、100 種類の分子記述子と Vd を説明変数とした予測モデルを構築した。次に最適な機械学習アルゴリズムを選択するために複数のアルゴリズムを検討した結果、Artificial Neural Networks(ANN)が最良の予測精度を示した($r=0.605$)。さらに、

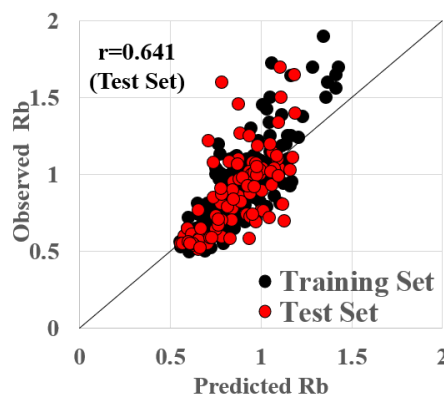


Fig. 1 Result of Rb prediction model

ANN モデルの計算コストを改善するため分子記述子の削減を検討したところ、6 種類の分子記述子と Vd を用いて構築した予測モデルにおいて $r=0.641$ と良好な結果を得た(Fig. 1)。この予測モデルは 100 種類の分子記述子と Vd で構築された予測モデルと比べて同等以上の精度を示した。本 Rb 予測モデルにおける Vd の影響を検討するために、Vd を除いた 6 種類の分子記述子のみで予測モデルを構築したところ、予測精度は低下した($r=0.506$)。Vd は Rb 予測モデルの精度に大きく影響したことから、Rb 予測における重要な因子であると推察された。

2. DeepSnap-DeepLearning 法および記述子法を組み合わせたラット CL 分類予測モデルの構築³⁾

創薬において、QSAR 解析を用いて予測される薬物動態パラメータの中には、精度が低いために予測モデルの実用が困難なものが存在する。このような薬物動態パラメータでは、新たな予測精度改善手法の開発が望まれている。近年、化合物の構造を画像としてとらえ、その画像を用いて DeepLearning により予測モデルを構築する DeepSnap-DeepLearning(DeepSnap-DL)法が開発され、毒性パラメータに対して高い予測精度を示した⁴⁾。そこで、予測が困難な薬物動態パラメータであるラット CL について、DeepSnap-DL 法を用いて高い予測精度を示す分類モデルの構築を試みた。本検討にはラット CL の実測値(平均: 2.0 L/hr/kg、

標準偏差: 2.1 L/hr/kg)を有する 1545 種類の社内化合物(訓練セット: 1236 化合物、テストセット: 309 化合物に層化抽出法で無作為に分割)を使用し、1 L/hr/kg を基準とした高 CL/低 CL 化合物分類モデルを構築した。特徴量には専用ソフトウェアによって算出した分子記述子を使用した。はじめに特徴量の重要度解析法である Permutation Importance により選出された上位 100 種類の記述子を用いて、統合的機械学習システムである DataRobot で選択された機械学習法である Random Forest により予測モデルを構築した(記述子法: Molecular Descriptor-Based Methods)。その結果、ROC 曲線下面積 (ROC AUC) は 0.883、Accuracy は 0.825 を示した(Table 1)。一方、DeepSnap-DL 法における ROC AUC および Accuracy は各々 0.905 および 0.832 であり(Table 1)、DeepSnap-DL 法は記述子法よりも高い予測精度を示した。

Table 1. The external test results for classification models

	n	ROC AUC	Accuracy
Molecular Descriptor-Based Methods	309	0.883	0.825
DeepSnap-DL Method	309	0.905	0.832
Ensemble Model	309	0.943	0.874
Consensus Model	221	-	0.959

n: Number of test compounds that can be evaluated.

さらに、本検討では記述子法および DeepSnap-DL 法の 2 つのモデルを組み合わせた新たな予測モデルを検証した(Fig. 2)。

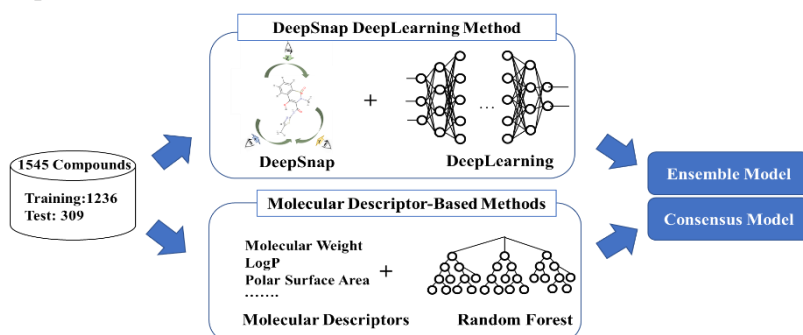


Fig. 2 Flowchart of the modeling process for rat CL prediction

それぞれの予測モデルの予測確率の平均値を新たな予測モデルの予測確率とした Ensemble Model において、ROC AUC および Accuracy は各々 0.943 および 0.874 となり、記述子法および DeepSnap-DL 法を単独で用いた場合よりも高い予測精度を示した (Table 1)。一方、記述子法および

DeepSnap-DL 法において予測結果が一致した結果のみを採用した Consensus Model を構築した。本モデルの評価可能な化合物数は 309 化合物から 221 化合物に減少するものの、その Accuracy は 0.959 と、これまでのラット CL に対する予測結果の中で最高精度を示した (Table 1)。

3. DeepSnap-DeepLearning 法および記述子法を組み合わせたラット CL 回帰予測モデルの構築⁵⁾

ラット CL の予測精度改善方法として記述子法および DeepSnap-DL 法の組み合わせによる分類モデルを構築した³⁾。しかし、分類モデルは大小二値の予測に留まる。一方、連続変数を返す回帰モデルは大きな実用上の利点を有するが、構築の難易度は高い。そこで、ラット CL を対象として新たな回帰モデル構築法の開発を試みた。本検討では分類モデルで使ったラット CL と同一の化合物データセットを使用した。また、分子記述子も分類モデル構築時と同様に算出した。はじめに分子記述子を用いた回帰モデルを構築するため、DataRobot によって選択されたアルゴリズムを使用した。その結果、Table 2 に示すように、決定係数(R^2)および二乗平均平方根誤差(RMSE)は各々 0.669 および 0.295 を示した(記述子法)。

Table 2. The external test results for regression models

	R^2	RMSE
Molecular Descriptor-Based Methods	0.669	0.295
Combination Model	0.769	0.247

次に、DeepSnap-DL 法によるラット CL 分類モデルを用いた各化合物の予測確率を算出し、回帰モデルの説明変数とした。すなわち、分類予測

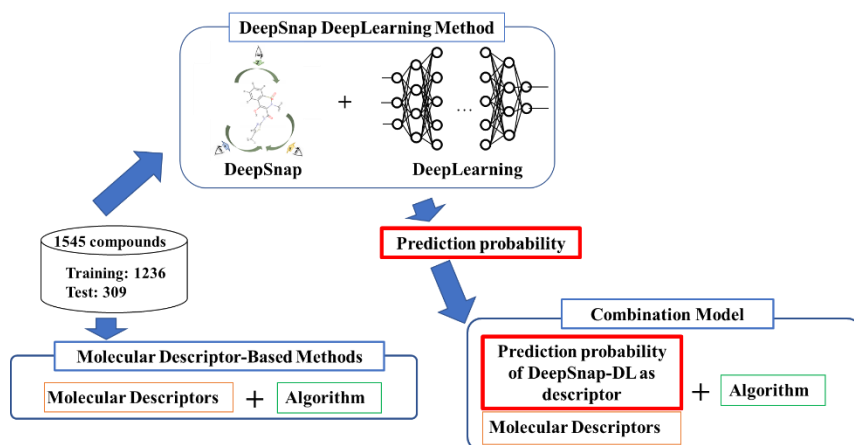


Fig. 3 Flowchart of the modeling process for rat CL prediction

確率と分子記述子を説明変数とした新たな回帰モデル(Combination Model)を構築した (Fig. 3)。これらの説明変数を、DataRobotによって選択された Random Forest を用いて解析した結果、 R^2 および RMSE は各々 0.769 および 0.247 となり、従来の記述子法に基づく QSAR 回帰モデルよりも高い予測精度のモデルの構築に成功した (Table 2)。

総括

本研究では、創薬および臨床上重要なパラメータである Rb の予測モデルを構築した。本予測モデルは分子記述子に加えて Vd を説明変数とする新規の予測モデルとして、分子記述子のみでは達成できなかった精度の予測に成功した。なお、本予測モデルはこれまでの報告の中で最大の化合物数を使用して構築されたことから、広い適用領域を有することが期待される。

さらに本研究では、分類モデルおよび回帰モデルで使用可能な記述子法および DeepSnap-DL 法の組み合わせによる新規の予測モデル構築手法を開発した。この構築手法はラット CL のみならず、他の薬物動態パラメータ、薬理活性、および毒性パラメータ等の予測にも適用可能である。応用範囲の広い本研究の結果が医薬品開発に有用な知見となることを期待したい。

参考文献

- 1) Mamada H., Iwamoto K., Nomura Y., Uesawa Y., *Mol Divers.*, **25**, 1261-1270 (2021)
- 2) Paixão P., Gouveia LF., Morais JAG., *Eur. J. Pharm. Sci.*, **36**, 544-554 (2009)
- 3) Mamada H., Nomura Y., Uesawa Y., *ACS Omega*, **6**, 23570-23577 (2021)
- 4) Uesawa Y., *Bioorg. Med. Chem. Lett.*, **28**, 3400–3403 (2018)
- 5) Mamada H., Nomura Y., Uesawa Y., *ACS Omega*, **7**, 17055-17062 (2022)