

**【目的】**創薬標的への強い活性を持つ新薬候補化合物の創成や、薬剤候補化合物の薬物動態パラメータの正確な推定は、医薬品研究開発の加速やコスト削減に直接繋がる重要課題である。本研究ではこれまであまり予測手法が確立されていなかった体内動態パラメータとして、薬剤の血液胎盤関門の透過性に着目し、これを予測するための学習手法を開発した。当初は深層ランク学習技術を活用することを検討していたが、ウェブアプリ等の形式でユーザーが迅速にかつ簡便に予測値を得られるようにするというアプリケーション実装を考慮した結果、勾配ブースティング木による予測手法構築を採用した。

**【方法】**血液胎盤関門透過性を表す指標である  $\log FM$  および  $CI$  を対象に、それぞれのデータセットを用いた予測モデルの構築を行った。 $\log FM$  データセットは *in vivo* 測定として臍帯血と母体血の血中濃度比 ( $F/M$  比) から得られたデータであり、Takaku らが文献から収集した 55 化合物の  $\log FM$  値を用いた。 $CI$  データセットは、*ex vivo* 実験であるヒト胎盤灌流実験によって化合物が母体側から胎児側へ移動する速度から求められる血液胎盤関門透過性の評価方法であり、コントロール化合物であるアンチピリン (antipyrene) に対するクリアランスの比率  $CI$  (clearance index) のデータである。 $CI$  データは Giaginis らが文献から収集した 88 化合物の  $CI$  値を用いた。予測モデルの構築では、化合物の特徴ベクトルとして *mordred* 記述子を用いた。各データセットは、それぞれ先行研究と同一の分割方法にて訓練データとテストデータに分割し、予測モデルの学習および検証に用いた。予測モデルの構築には勾配ブースティング木の実装の一つである *LightGBM* を用い、*optuna* の *LightGBM Tuner* によってパラメータチューニングを行った。評価指標は決定係数  $R^2$  と根平均二乗誤差 *RMSE* を用いた。また、構築した *LightGBM* による予測モデルを用いて、テキストボックスに入力された化合物の *SMILES* 文字列から  $CI$  と  $\log FM$  の予測値を計算するウェブサービスを構築した。フロントエンドは *Vue.js* フレームワークを用い、バックエンドは *Python* と *Flask* を用いてシングルページアプリケーションとして構築した。あわせて [https://pbpredictor.net/api/predict?smiles=\[str\]](https://pbpredictor.net/api/predict?smiles=[str]) の *URI* によって *SMILES* 文字列から  $\log FM$  と  $CI$  の予測値をリクエスト可能な *REST API* を実装した。

**【結果】**提案手法の *LightGBM* による予測は、 $\log FM$  と  $CI$  の双方で従来手法の予測精度を上回る結果となった。特に、従来手法ではテストデータの精度が訓練データに比べて良くなかったが、提案手法ではテストデータでも良好な予測性能を示しており、汎化性能が良いことが示唆された。構築したウェブサービスは <https://pbpredictor.net> からアクセス・利用が可能である。

予測ウェブサービス *PBPredictor* の概観 <https://pbpredictor.net>

No.	SMILES	$\log FM$	$CI$
1	C1CC1NC2=C3C=NC(=N2)N(C=C3)C(=O)C4=O	-0.09	0.48
2	C1C(O)C(S1)O(NC=O)N=C2=O	-0.05	0.38
3	CC1=CC=C(C=C1)C(=O)C2=CC=CC=C2	0.91	0.62
4	IN(=O)S(=O)(=O)N		
5	CC(=O)NC1=CC=C(O)C=C1	0.09	0.53

© Ohue Laboratory, Tokyo Institute of Technology