

GBDTによる化合物の血液胎盤関門透過性予測

大上 雅史^{1,a)} 古井 海里¹

概要: 母体に投与された薬剤が胎児にどの程度移行するかは、胎盤にある血液胎盤関門の透過性によって評価される。化合物の血液胎盤関門透過性を知ることは安全な薬剤を開発する上で重要であり、また妊婦の治療における指針にもなりうる。本研究では、勾配ブースティング木 (GBDT) を用い、比較的軽量の計算で血液胎盤関門透過性を予測することを試みた。2種類の公開データ (出産時の臍帯血および母体血中の薬物濃度測定で得られる *in vivo* データ (logFM)、およびヒト胎盤灌流実験で得られる *ex vivo* データ (CI)) からそれぞれの回帰モデルを構築し、logFM および CI それぞれにおいて従来手法の精度を超える予測を達成した。また、構築した回帰モデルによる予測を提供するウェブサービスを構築した。

キーワード: 血液胎盤関門透過性、教師あり学習、勾配ブースティング木 (GBDT)、胎児母体血中濃度比 (logFM)、clearance index (CI)

1. 序論

母体に投与された薬剤が胎児にどの程度移行するかは、胎盤にある血液胎盤関門をどの程度透過するかという指標で評価される。薬剤候補となる化合物の血液胎盤関門透過性を知ることは安全な薬剤を開発する上できわめて重要であり、また妊婦の治療における指針にもなりうる [1]。しかし、ヒトの胎盤は他の動物モデルでは再現されない種特異的性質を持っており、ヒトの組織モデルを用いた評価が必要となっている [2]。

化合物の血液胎盤関門透過性の評価には、出産時の臍帯血 (胎児血) にどの程度化合物が含まれるかを調べた胎児-母体血中濃度比 (F/M 比) や、ヒト臍帯血を用いた胎盤灌流モデルによる *ex vivo* 実験が用いられる。F/M 比は *in vivo* データとして血液胎盤関門透過性の直接的な指標となるが、原理的に安全であるとわかっている薬剤のデータしか得られないという制限がある [3]。一方の胎盤灌流モデルによる *ex vivo* 実験であればどのような化合物でも測定できるものの、あくまで間接的な指標であることに注意が必要である [4]。

本研究では、現在までに得られている F/M 比のデータや胎盤灌流モデルによる実験データから、血液胎盤関門透過性を予測することを試みる。これまでに機械学習による血液胎盤関門透過性予測についてはいくつかの事例が存在す

る [5], [6], [7], [8]。しかし、いずれも予測精度は不十分であり、また予測手法を利用できるソフトウェア等も存在していなかった。そのため、本研究では機械学習による予測を広く利用できるように比較的軽量の計算で高精度な予測が可能な勾配ブースティング木 (Gradient Boosting Decision Tree, GBDT) に着目した。GBDT を用いて、2種類の公開データ (出産時の臍帯血および母体血中の薬物濃度測定で得られる *in vivo* データ (logFM)、およびヒト胎盤灌流実験で得られる *ex vivo* データ (CI)) からそれぞれの回帰モデルを構築し、より高精度な予測を行い、さらに構築した回帰モデルによる予測を実際に提供するウェブサービスを構築することを目的とする。

2. 手法

2.1 データセット

本研究では、血液胎盤関門透過性を表す以下の2つの指標を対象に、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。

logFM データセット ヒトにおける化合物の胎盤移行を直接測定することは困難であるが、*in vivo* 測定として臍帯血と母体血の血中濃度比 (F/M 比) のデータが存在する (文献 [3] 等)。logFM は F/M 比の常用対数をとった値である。本研究では Takaku らが文献から収集した 55 化合物の logFM 値を用いた [5]。このデータセットでは、Takaku ら [5] が多重線形回帰により、また Wang ら [6] が Lasso 回帰による予測手法を提案している。本研究では、Wang らの予測を比較対象として取り上げた。

¹ 東京工業大学 情報理工学院 情報工学系
Department of Computer Science, School of Computing,
Tokyo Institute of Technology

^{a)} ohue@c.titech.ac.jp

CI データセット *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（クリアランス）を求めることで血液胎盤関門透過性を評価できる。コントロール化合物であるアンチピリン（antipyrine）に対するクリアランスの比率を CI (clearance index) といひ、CI によって化合物の胎盤移行が評価される [9]。胎盤灌流実験は介入を必要としないため、logFM に比べてデータが取りやすいという利点がある。本研究では Giaginis らが文献から収集した 88 化合物の CI 値を用いた [7]。このデータセットでは、Giaginis ら [7] および Zhang ら [8] が Partial Least Square を用いた予測手法を提案している。本研究では、Zhang らの予測を比較対象として取り上げた。

2.2 予測モデルの構築

化合物の特徴ベクトルとして、mordred 記述子 [10] を用いた。mordred は RDKit を介して 1,613 種類の 2D 記述子と 213 種類の 3D 記述子が計算できる。3D 記述子の計算に必要な立体配座生成には、RDKit の ETKDGv2[11] を用いた。

各データセットは、2.1 節で示した比較対象文献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た。回帰モデルの構築には勾配ブースティング木 (GBDT) の実装の 1 つである LightGBM[12] を用いた。また、パラメータチューニングは訓練データの Leave-One-Out 交差検定 (LOOCV) の結果に基づき、optuna の LightGBM Tuner[13] を用いて行った。ただし事前に確認したところ、特に logFM の予測で過学習が起きており、訓練データの LOOCV は決定係数 $R^2 = 1$ となりテストデータの R^2 は極端に悪化するという結果であった。そのため、Lasso 回帰による特徴選択を訓練データの LOOCV で事前に行い、計 5 個の記述子を選択して LightGBM の入力とした。

2.3 評価指標

決定係数 R^2 と根平均二乗誤差 RMSE を用いた。optuna の最適化指標には LOOCV での RMSE 値を用いた。

2.4 ウェブサービス実装

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 LogFM 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[14] フレームワークを用い、バックエンドは Python と Flask[15] 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 logFM と CI の予測値をリクエスト可能な REST API を実装した。

表 1 logFM の予測結果

logFM 予測	提案手法 (LightGBM)		Wang2020[6]	
	R^2	RMSE	R^2	RMSE
訓練 (LOOCV)	0.716	0.181	0.633	0.206
テストデータ	0.720	0.191	0.557	0.241

表 2 CI の予測結果

CI 予測	提案手法 (LightGBM)		Zhang2015[8]	
	R^2	RMSE	R^2	RMSE
訓練 (LOOCV)	0.175	0.269	(0.906)	(0.090)
テストデータ	0.624	0.158	0.171	0.234

※ Zhang2015 [8] では LOOCV ではなく訓練データ全体を用いて学習したときの R^2 と RMSE が報告されているため、その値を () 付きで示した。

表 3 提案手法で選択された 5 つの記述子

logFM	AATSC1c, MAXsCH3, ZMIC1, Lipinski, PEOE_VSA6
CI	ATSC1se, AATSC0c, GATS3v, PPSA3, DPSA3

3. 結果と考察

3.1 血液胎盤関門透過性の予測結果

logFM 値の予測結果を表 1 に、CI 値の予測結果を表 2 に示す。提案手法の LightGBM による予測は、logFM 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る。

3.2 選択された記述子の考察

表 3 に提案手法で選択された記述子を示した。一般に、胎盤を透過しやすい薬剤は、分子量が小さく (< 500)、脂溶性で、血漿タンパク質結合率が低い、という傾向が知られている [16]。表 3 に示された記述子のうち、ZMIC1 と Lipinski 以外は空間電荷に関わる記述子であり、脂溶性との関わりが考えらる。また、ZMIC1 は原子間結合の分岐やサイズと関係しており、分子量との相関が考えられる。Lipinski はリピンスキー則 (rule of five) を満たすか否かを表す記述子であり、分子量や脂溶性と直接関わっている。AATSC1c と ZMIC1 については従来手法 [6] でも選択された記述子であり、より血液胎盤関門透過性との関係が期待される。

なお、血液脳関門透過性についても類似する法則性が示唆されているが [17]、血液脳関門透過性の予測値 (ADMET Predictor[18] で計算) と logFM や CI との強い相関は得られなかった。

3.3 ウェブサービスの利用例

構築したウェブサービスの概観を図 1 に示す。テキストボックスに SMILES 文字列を 1 化合物 1 行で入力し、



図 1 予測ウェブサービス PB Predictor の概観
<https://pbpredictor.net>

submit ボタンを押すと logFM と CI の予測値が表示される。なお、SMILES として不適合な文字列が入力された場合は INVALID SMILES と表示されるようにした。

また、API リクエストの例として、[https://pbpredictor.net/api/predict?smiles=CC\(=O\)NC1=CC=C\(C=C1\)O](https://pbpredictor.net/api/predict?smiles=CC(=O)NC1=CC=C(C=C1)O) のように URL を指定すると、{"logFM":0.0919, "CI":0.5330, "smiles":"CC(=O)NC1=CC=C(C=C1)O"} のように json 形式で値が返ってくるようになっている。

4. 結論

本研究は、勾配ブースティング木 (GBDT) を用いて化合物の血液胎盤関門透過性を予測する手法を提案した。GBDT の枠組みの 1 つである LightGBM を用いて予測モデルを構築した結果、従来手法と比較して良好な予測性能を示した。提案手法はテストデータでも予測精度が低下しなかったことから、汎化性能もある程度担保できているものと考えられる。LightGBM は計算が軽量であることも利点の 1 つであり、ウェブサーバを介して予測結果を即座に返すようなサービスにも適している。本研究では PB Predictor と名付けた血液胎盤関門透過性予測ウェブサービスを試作し、公開した。

本研究では、記述子の組合せ方や他の記述子の利用など、予測手法構築の上で検討が及んでいない点が多々存在する。また、logFM と CI の両方を同時に考慮するマルチラベル学習の検討や、血液脳関門透過性や皮膚透過性など他の物性指標との関係、催奇形性や胎児毒性などの毒性との関係に関する考察も不十分である。さらに、化合物の部分構造に起因する血液胎盤関門透過性への影響を示唆でき

れば、薬剤設計においてより安全な部分構造の提案なども可能になると考えられ、これらは今後の課題である。

謝辞 本研究は、JSPS 科研費 (20H04280)、JST ACT-X (JPMJAX20A3)、公益財団法人 上原記念生命科学財団の支援を受けて行われた。

参考文献

- [1] Myren M, *et al.* The human placenta—An alternative for studying foetal exposure. *Toxicol in Vitro*, 21(7):1332–1340, 2007.
- [2] Giaginis C, *et al.* Current toxicological aspects on drug and chemical transport and metabolism across the human placental barrier. *Exp Opin Drug Metab Toxicol*, 8(10):1263–1275, 2012.
- [3] Fukata H, *et al.* Necessity to measure PCBs and organochlorine pesticide concentrations in human umbilical cords for fetal exposure assessment. *Environ Health Perspect*, 113(3):297–303, 2005.
- [4] Hutson JR, *et al.* The human placental perfusion model: a systematic review and development of a model to predict *in vivo* transfer of therapeutic drugs. *Clin Pharmacol Ther*, 90(1):67–76, 2011.
- [5] Takaku T, *et al.* Quantitative structure-activity relationship model for the fetal-maternal blood concentration ratio of chemicals in humans. *Biol Pharm Bull*, 38(6):930–934, 2015.
- [6] Wang C-C, *et al.* Prediction of human fetal-maternal blood concentration ratio of chemicals. *PeerJ*, 8:e9562, 2020.
- [7] Giaginis C, *et al.* Application of quantitative structure-activity relationships for modeling drug and chemical transport across the human placenta barrier: a multivariate data analysis approach. *J Appl Toxicol*, 29(8):724–733, 2009.
- [8] Zhang Y-H, *et al.* Prediction of placental barrier permeability: a model based on partial least squares variable selection procedure. *Molecules*, 20(5):8270–8286, 2015.
- [9] Smith JA, *et al.* Utilization of an ex vivo human placental perfusion model to predict potential fetal exposure to carboplatin during pregnancy. *Am J Obstet Gynecol*, 210(3):275.e1–9, 2014.
- [10] Moriwaki H, *et al.* Mordred: a molecular descriptor calculator. *J Cheminform*, 10:4, 2018.
- [11] Riniker S & Landrum GA. Better informed distance geometry: Using what we know to improve conformation generation. *J Chem Inf Model*, 55:2562–2574, 2015.
- [12] Ke G, *et al.* LightGBM: a highly efficient gradient boosting decision tree. *Adv Neural Inf Process Syst*, 30:3146–3154, 2017.
- [13] LightGBM Tuner: New Optuna Integration for Hyperparameter Optimization, <https://link.medium.com/JxrfBJ5tVdb>
- [14] Vue.js, <https://vuejs.org>
- [15] Aslam FA, *et al.* Efficient way of web development using python and flask. *Int J Adv Res Comput Sci*, 6(2):16–19, 2015.
- [16] Joshi AA, *et al.* Placental ABC transporters: biological impact and pharmaceutical significance. *Pharm Res*, 33(12):2847–2878, 2016.
- [17] Abbott NJ, *et al.* Structure and function of the blood-brain barrier. *Neurobiol Dis*, 37(1):13–25, 2010.
- [18] ADMET Predictor version 10.0, <https://www.simulations-plus.com/software/>