

## 抗がん活性予測を利用した化合物提案 AI システムの開発

(早大先進理工<sup>1</sup>・早大理工総研<sup>2</sup>) ○増田 幹太<sup>1</sup>・金子 武史<sup>1</sup>・中嶋 裕也<sup>1</sup>・町田 光史<sup>1</sup>・神平 梨絵<sup>1</sup>・中尾 洋一<sup>1,2</sup>・清野 淳司<sup>1,2</sup>

Development of an AI system for proposing compounds using anti-cancer activity prediction.

(<sup>1</sup>*School of advanced Science and Engineering, Waseda University*, <sup>2</sup>*Waseda Research Institute for Science and Engineering*) ○Kanta Masuda<sup>1</sup>, Takeshi Kaneko<sup>1</sup>, Yuya Nakajima<sup>2</sup>, Koshi Machida<sup>1</sup>, Rie Kamihira<sup>1</sup>, Yoichi Nakao<sup>1,2</sup>, Junji Seino<sup>1,2</sup>

We have developed an AI system that predicts biological activities with high accuracy for a wide range of compounds and bioactive species or cell lines. In this study, we extended the system to extract safe compounds by incorporating the prediction of safety-related properties, implicitly addressing absorption, distribution, metabolism, excretion, and toxicity (ADMET).

**Keywords** : Machine Learning; Human Cancer Cell; Anti-Cancer Action of Compounds

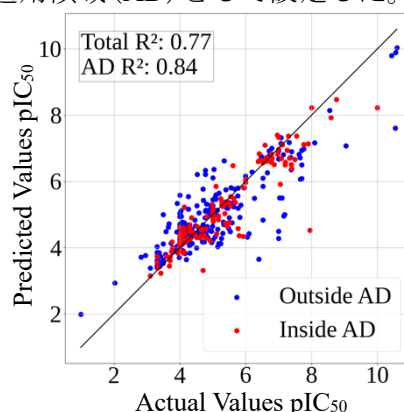
【緒言】AI 技術と創薬ビッグデータの活用により、効能・効果の高い抗がん剤開発の加速が期待されている。これまでに我々は、広範な化合物や生物活性種/細胞腫に対する高精度の生物活性予測 AI システムを開発してきた<sup>1</sup>。当システムを用いた多角的な予測に基づき、類似の性質を持つ化合物を探索できる。本研究では、毒性・安全性に関する性質の予測を追加することで、吸収・分布・代謝・排泄・毒性 (ADMET) を暗に考慮した安全な化合物を抽出するシステムに拡張した。

【予測モデルの構築】本研究では化合物の毒性の指標として、正常細胞とがん細胞に対する阻害能の相対値を使用した。阻害能として 50%阻害濃度 (IC<sub>50</sub>) を用いた。また安全性の指標としてヒト経口バイオアベイラビリティ (HOB) を採用した。IC<sub>50</sub> のデータは、アメリカ国立がん研究所 (NCI) のヒトがん細胞株 60 種に対する薬剤感受性データ<sup>2</sup>を用いた。また正常細胞としては線維芽細胞の WI-38 を選定し、ChEMBL<sup>3</sup> からデータを取得した。HOB データを admetSAR<sup>4</sup> から収集し、HOB の値が 50 以上か否かを分類するタスクを設定した。両方の指標の予測において、構造記述子には RDKit ライブラリ中の分子フィンガープリントおよび 2D/3D 記述子を使用し、機械学習モデルには AutoGluon ライブラリの TabularPredictor を用いた。さらに、k 近傍法を用いて学習データの 50% の距離の範囲内をモデル適用領域 (AD) として設定した。

【結果】正常細胞での pIC<sub>50</sub> の予測結果を図 1 に示す。テストデータ全体で決定係数 (R<sup>2</sup>) が 0.73、AD 内のデータで 0.84 と高い精度で予測できることが確認された。また HOB 予測 (表 1) についても、テストデータ全体の精度が 0.76、AD 内で 0.79 と高い精度で予測できた。発表当日は、これらの予測を基に *in silico* で行った化合物提案の結果も示す予定である。

**Table 1.** Confusion matrix for the classification results of HOB data within the AD.

|        |          | Predicted |          |
|--------|----------|-----------|----------|
|        |          | 50 ≤ HOB  | HOB < 50 |
| Actual | 50 ≤ HOB | 66        | 8        |
|        | HOB < 50 | 15        | 19       |



**Figure 1.** Scatter plot in pIC<sub>50</sub> predictions for WI-38 cell.

1) 清野淳司ら、特願 2024-091295, **2024**. 2) R. Shoemaker, *Nat Rev Cancer*. **2006**, 6, 813. 3) M. Davies, *Nucleic Acids Res*. **2015**, 43, 612-620. 4) M. Wei, et al. *J. Cheminform*. **2022**, 14, 1.