

## 機械学習を用いたヒトがん細胞に対する化合物の抗がん活性の予測解析

(早大先進理工<sup>1</sup>・早大理工総研<sup>2</sup>) ○金子 武史<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>

Prediction of anticancer activity by chemical regents on human cancer cells using machine learning (<sup>1</sup>*School of advanced Science and Engineering, Waseda University*, <sup>2</sup>*Waseda Research Institute for Science and Engineering*) ○Takeshi Kaneko<sup>1</sup>, Koshi Machida<sup>1</sup>, Rie Kamihira<sup>1</sup>, Yoichi Nakao<sup>1,2</sup>, Junji Seino<sup>1,2</sup>

In search of natural products, it is also important to discover the unknown biological activity of known compounds. However, development of the predictions of the activity from chemical structures remains challenging. In this study, we focused on anticancer activity and developed a model to predict the cytotoxic activity of compounds using machine learning and a huge amount of drug sensitivity data in NCI-60.

**Keywords :** *Machine Learning; Human Cancer Cell ; Anticancer Activities of Compounds*

【緒言】天然化合物の探索研究では、新規化合物の発見とともに、既知化合物の新規生物活性を見出すことも同様に重要である。多様な構造に適用可能な化学構造から生物活性を予測する手法の開発研究は盛んになりつつあるが、広く実用化される段階には至ってはいない。そこで本研究では簡便に生物活性予測が可能となる手法の確立を目指し、がん細胞パネル試験による抗がん活性評価結果と化学構造の関係性に着目し、機械学習を用いて化学構造から抗がん活性を予測するモデルを開発した。

【抗がん活性の予測】本研究では、アメリカ国立がん研究所（NCI）で確立されたヒトがん細胞株60種に対する抗がん剤スクリーニング用パネル内の薬剤感受性データを活用した<sup>1)</sup>。予測値として50%増殖阻害濃度値（GI50）を採用した。約25,000個の化合物を用い、各細胞株に対してモデルを構築した。機械学習手法としてランダムフォレスト回帰を、記述子としてRDKitライブラリに含まれる分子フィンガープリントおよび2次元分子記述子を用いた<sup>2)</sup>。ハイパーパラメータはOptunaを利用して最適化し、五分割交差検証により予測性能を評価した。

NCI-H23 細胞株に対する予測では、平均絶対誤差（MAE）が0.338、2乗平均平方根誤差（RMSE）が0.504、決定係数（R<sup>2</sup>）が0.835となり（Fig. 1）、多くの化合物に対して高い精度でGI50を予測できることが確認された。また、59種の細胞株に対するR<sup>2</sup>値が0.7以上となり（Fig. 2）、細胞種によらない高い予測精度が得られた。

1) [https://ntp.cancer.gov/discovery\\_development/nci-60](https://ntp.cancer.gov/discovery_development/nci-60).

2) <https://www.rdkit.org>.

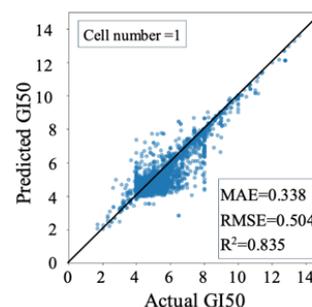


Figure 1. Scatter plot in GI50 predictions for NCI-H23 cell.

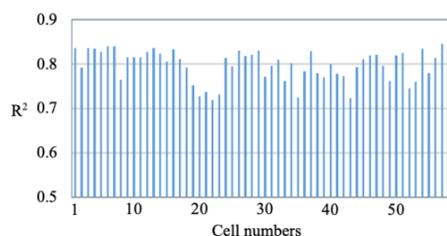


Figure 2. Results of R<sup>2</sup> values in 59 cells.