

分子軌道エネルギーを説明変数とした機械学習による薬効予測

寺前 裕之^{a*}, 三浦 優太^a, 色摩 光一^b, 玄美 燕^b, 高山 淳^b, 岡崎 真理^b, 坂本 武史^b^a城西大学理学部, 〒350-0295 埼玉県坂戸市けやき台1-1^b城西大学薬学部, 〒350-0295 埼玉県坂戸市けやき台1-1
teramae@gmail.com

(Received: July 27, 2024; Accepted for publication: November 8, 2024; Online publication: January 15, 2025)

We constructed a mathematical model to predict the 2,2-diphenyl-1-picrylhydrazyl (DPPH) free radical scavenging capacity (IC₅₀) for recently synthesized ferulic acid derivatives by machine learning with molecular orbital energy as an explanatory variable and IC₅₀ as an objective variable. We compared 96 regression models including xgbLinear and neuralnet included in R/caret package. We were able to construct IC₅₀ prediction models for these new ferulic acids by using xgbLinear, M5, ppr, and neuralnet as regression methods.

Keywords DPPH free radical scavenging capacity, pharmacological activity, ferulic acid, IC₅₀, machine learning

1 はじめに

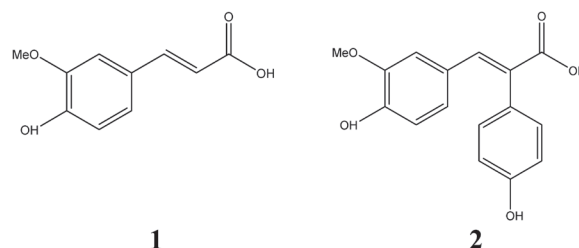
近年、機械学習が注目を集めており、化学分野への応用の可能性も広がってきた。我々は分子軌道法により求められた分子軌道エネルギーを説明変数として使う事で様々な化学事象の解明に使用できるのではないかと考え、一連の研究を行ってきた。

我々は、2008年にベンゾジアゼピンおよびチエノジアゼピン系の抗不安薬について、抗不安性と抗痙攣性の強さとnext-HOMOの軌道エネルギー値の間に相関があることを初めて見出した [1]。その後の発表でフェルラ酸 (FA) の抗酸化作用に対して、DPPH フリーラジカル消去濃度 (IC₅₀) の置換基効果を説明できることを示した [2–4]。ただし、ただ一つの軌道エネルギー値に対して強い相関関係を得ることは不可能であったため、複数の軌道エネルギー値と IC₅₀ との関係 Random Forest 回帰法により機械学習を試みたところ、訓練データのみではあるが、決定係数 0.89 という強い相関関係が見いだされ [3]、分子軌道エネルギーと機械学習のみで IC₅₀ 値の予測が可能となった。

2022年度の発表ではフェルラ酸誘導体の IC₅₀ 値をブースティングやニューラルネットなど Random Forest 以外の回帰法を用いて機械学習を行った結果について報告した。フェルラ酸についてはラジカル状態を考えて UHF 法による分子軌道エネルギーを使用する必要もなく、元々のフェルラ酸誘導体の分子軌道エネルギーのみにより予

測を行えることも示した [4]。

最近の実験において、従来のフェルラ酸 **1** に対して置換基としてフェニル基を導入した **2** と一連の 29 種類の置換体の合成が行われ、これらのラジカル消去能の予測には従来のフェルラ酸のみで構成した数理モデルでは不十分なことが判明した。そこで本研究ではこれらの新たなフェルラ酸誘導体のデータを使用した新しい数理モデルの構築を試みた。



2 計算方法

分子軌道計算プログラム Gaussian16 Rev.B.01 を使用し RHF/6-31G(d,p) レベルで構造最適化を行った [5]。最適化された構造に対して振動数計算を行い、少なくとも局所安定構造であることを確認した。

機械学習計算には R の caret パッケージ [6] を使用し分子軌道計算により得られた軌道エネルギー値を説明変数

Table 1. Coefficient of determination R^2 for test data and R^2 for test data and training data at the number of orbitals giving the minimum RMS error (RMSE) between the predicted and experimental values of the test data within the range of 2 to 80 molecular orbitals for 14 different regression methods.

Method	# of orbitals	R^2 (Test set)	R^2 (Total)	RMSE
SBC	36	0.980	0.992	0.164
mlp	40	0.979	0.993	0.171
qrnn	62	0.965	0.991	0.183
monmlp	38	0.961	0.991	0.186
mlpWeightDecay	42	0.976	0.986	0.209
xgbLinear	14	0.959	0.986	0.242
M5	2	0.953	0.831	0.272
ppr	4	0.927	0.868	0.272
xgbTree	66	0.942	0.981	0.277
neuralnet	14	0.942	0.980	0.284
kknn	26	0.973	0.866	0.303
svmLinear3	18	0.881	0.738	0.354
HYFIS	20	0.869	0.924	0.357
gamboost	14	0.828	0.756	0.391

として、最大80軌道までを、また96種類の回帰法を使用して解析を試みた。トレーニングデータは24分子、未知データとして扱うテストデータは8分子とした。データの分割は、乱数を使用して行った。Validationは3-fold-cross-validationを使用した。回帰法はcaretパッケージに含まれるものを全て試みたが94種類について解が得られた。テストデータに対する予測値と実験値の差のRMSE値が最小になる時の分子軌道エネルギーの数を説明変数の最適な数とした。

3 結果と考察

Table.1に2~80個の分子軌道数で、テストデータも含めた場合の決定係数 R^2 、トレーニングデータの決定係数の値 R^2 、テストデータの予測値と実験値との誤差であるRMSEの最小値、をRMSEの小さい順に14種類の回帰法について示した。

RMSEの最小値は36軌道のエネルギーを説明変数として使用したSBCが0.164と最も良いスコアになっている。その次はmlpの0.171であるが、これは軌道数が40の時の値である。さらに3番目はqrnnで0.183であり軌道数は62である。4番目と5番目はそれぞれmonmlpとmlpWeightDecayであるが、いずれもRMSEの値は小さいが軌道数は38と42となっていて、ここまでの5種類については軌道数が全て大きい値となっている。6-8番目についてはxgbLinear, M5, pprで軌道数は14, 2, 4となっていて比較的小きな値である。9番目はxgbTreeでxgbLinearに比べて66と大きな値となっている。10番目はneuralnetで軌道数は14である。ちなみに我々の研究

でneuralnetが良いスコアを示したのはこれが初めてである。11-14番目はkknn, svmLinear3, HYFIS, gamboostとなっていて軌道数はそれぞれ26, 18, 20, 14である。分子軌道エネルギーの数という物理的意味からは大きい軌道数はやや疑問であると考えているため、軌道数が比較的少なかったxgbLinear, M5, ppr, neuralnet, kknn, svmLinear3, HYFIS, gamboost.の8種類についてさらに詳しく見ていく。

Figure 1に最適な軌道数における、xgbLinear, M5, ppr, neuralnet, kknn, svmLinear3, HYFIS, gamboostの実験値と予測値の相関関係を表す散布図を示した。全体的な決定係数においてはxgbLinearとneuralnetが0.986と0.980と大きな値となっており、M5とpprの0.831と0.868よりは良い相関関係を示している。ただし、M5とpprにおいては軌道数が2と4であって、より少ない説明変数で最適なスコアを得ることができている。kknnについては軌道数が26と少し大きく、svmLinear3, HYFIS, gamboostの3種類についてはRMSE値がやや大きくなっている。

なぜ、この方法により予測が可能になるかについてであるが、どの軌道が反応と関与しているのかを機械学習により決定していることになるので、基本的にはフロンティア軌道理論 [7]の自然な拡張になっているためと考えられる。これについてはさらなる検討が必要であろう。

本論文ではTable 1またFigure 1においても他の研究で得られたものよりも決定係数が比較的小さく、軌道数が少ない結果が少数しか得られなかったため、今後も異なる観点からの検討が必要と思われる。

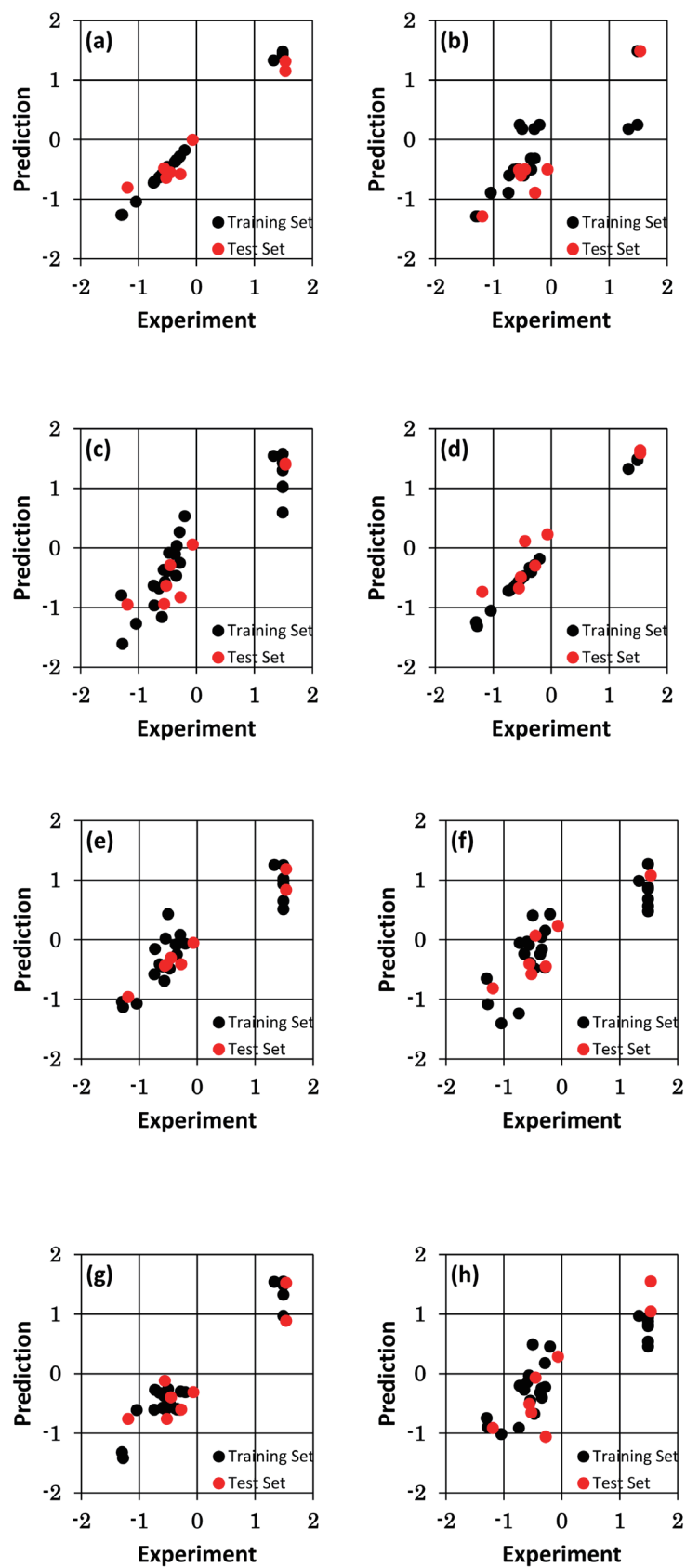


Figure 1. Correlation between experimental and predicted values with (a) xgbLinear, (b) M5, (c) ppr, (d) neuralnet, (e) kkn, (f) svmLinear3, (g) HYFIS, and (h) gamboost.

4 結論

本研究では近年新たに合成されたフェルラ酸誘導体について、分子軌道エネルギーを説明変数としてDPPHフリーラジカル消去能(IC₅₀)を目的変数とした機械学習を行うことで、IC₅₀を予測する数理モデルを構築した。回帰法としてはxgbLinearやneuralnetを始めとした96種類を比較した。回帰法としてxgbLinear, M5, ppr, neuralnetを使用することにより、これらの新規フェルラ酸類のIC₅₀を予測する数理モデルを構築することが可能となったと考えられる。

参考文献

- [1] H. Teramae, K. Ohtawara, T. Ishimoto, U. Nagashima, *Bull. Chem. Soc. Jpn.*, **81**, 1094 (2008). <https://doi.org/10.1246/bcsj.81.1094>
- [2] H. Teramae, M. Xuan, T. Yamashita, J. Takayama, M. Okazaki, T. Sakamoto, *J. Comput. Chem. Jpn.*, **17**, 150 (2018). <https://doi.org/10.2477/jccj.2018-0018>
- [3] H. Teramae, M. Xuan, T. Yamashita, J. Takayama, M. Okazaki, T. Sakamoto, *J. Comput. Chem. Jpn.*, **18**, 211 (2019). <https://doi.org/10.2477/jccj.2019-0034>
- [4] H. Teramae, M. Xuan, J. Takayama, M. Okazaki, T. Sakamoto, T. Sakamoto, *J. Comput. Chem. Jpn.*, **21**, 103 (2022). <https://doi.org/10.2477/jccj.2023-0001>
- [5] Gaussian 16, Revision B.01, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox, Gaussian, Inc., Wallingford CT, 2016.
- [6] <https://cran.r-project.org/web/packages/caret/caret.pdf>
- [7] K. Fukui, *Chemical Reaction and Orbital of Electron*, Maruzen (1976) K. Fukui, *Theory of Orientation and Stereoselection*, Springer (1970).