



## Article

# Prediction and Chemical Interpretation of Singlet-Oxygen-Scavenging Activity of Small Molecule Compounds by Using Machine Learning

Taiki Fujimoto and Hiroaki Gotoh \*

Department of Chemistry and Life Science, Yokohama National University, 79-5 Tokiwadai, Hodogaya-ku, Yokohama 240-8501, Japan; fujimoto-taiki-gc@ynu.jp

\* Correspondence: gotoh-hiroaki-yw@ynu.ac.jp

**Abstract:** A chemically explainable machine learning model was constructed with a small dataset to quantitatively predict the singlet-oxygen-scavenging ability. In this model, ensemble learning based on decision trees resulted in high accuracy. For explanatory variables, molecular descriptors by computational chemistry and Morgan fingerprints were used for achieving high accuracy and simple prediction. The singlet-oxygen-scavenging mechanism was explained by the feature importance obtained from machine learning outputs. The results are consistent with conventional chemical knowledge. The use of machine learning and reduction in the number of measurements for screening high-antioxidant-capacity compounds can considerably improve prediction accuracy and efficiency.

**Keywords:** machine learning; antioxidant; singlet oxygen; feature importance; interpretability; carotenoid



**Citation:** Fujimoto, T.; Gotoh, H. Prediction and Chemical Interpretation of Singlet-Oxygen-Scavenging Activity of Small Molecule Compounds by Using Machine Learning. *Antioxidants* **2021**, *10*, 1751. <https://doi.org/10.3390/antiox10111751>

Academic Editor: Riccardo Amorati

Received: 11 October 2021

Accepted: 28 October 2021

Published: 1 November 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Oxygen is essential for human life. However, in the human body, some amount of oxygen exists as reactive oxygen species, which contributes to the immunity mechanism. However, aging and carcinogenesis are also attributed to reactive oxygen species because of their high reactivity [1]. Antioxidants are substances that remove reactive oxygen species in the body, and the ability to scavenge reactive oxygen species is called antioxidant capacity. Therefore, dietary intake of antioxidants through vegetables, fruits, crustaceans, and other foods is recommended [2].

The method for measuring the scavenging activity depends on the target reactive oxygen species. Furthermore, the use of only one method for determining the antioxidant capacity of compounds and foods may not provide reliable results. The oxygen radical absorbance capacity (ORAC) assay [3,4] was developed as an antioxidant capacity assay for peroxy radicals, which are commonly used as an indicator of the antioxidant capacity of foods and compounds [5]. Phenolic antioxidants exhibit high radical-scavenging capacity. However, carotenoids exhibit low peroxy radical-scavenging ability and high singlet-oxygen-scavenging ability [6]. The singlet-oxygen-scavenging ability is critical for peroxy radical scavenging [7]. The singlet oxygen absorbance capacity (SOAC) assay [8] was developed to measure the singlet oxygen. Research on the effects of foods and compounds on singlet oxygen is ongoing. The Antioxidant and Functional Research Group [9], established in Japan to promote research on antioxidant and functional properties of foods, named the general antioxidant capacity of foods antioxidant unit and defined it as the sum of ORAC and SOAC measurements.

The analysis of antioxidant capacity is time consuming and expensive. To screen compounds from candidates, the development of a simple and rapid method for measuring the antioxidant capacity of compounds is critical. Machine learning was widely used in various fields, such as industry, medicine, and engineering, for efficient research, and

diagnosis [10,11]. In the chemical field, the use of machine learning is promoted in domains such as toxicity determination [12], retrosynthesis [13], and prediction of the antioxidant capacity of polyphenols [14]. Machine learning can efficiently evaluate physical properties and provide details for searching the structure of new substances. In chemical studies, preparing a large amount of data for training is difficult, and machine learning should be performed with a small amount of data. Therefore, a method to obtain accurate predictions from a small amount of data is necessary.

In this study, a regression model that predicts singlet-oxygen-scavenging capacity as an objective variable was constructed by using the molecular structure as an explanatory variable. The numerical values obtained from computational chemistry and Morgan fingerprinting [15] were used as the input data. To obtain the interpretability of the machine learning model, regression coefficients were obtained from the linear model, and the feature importance was obtained from the prediction model based on the decision tree [16]. To verify the interpretability of the model, we attempted to explain the reaction mechanism [17] of singlet oxygen scavenging from the feature importance.

## 2. Experiment

### 2.1. Preparing Dataset

The singlet-oxygen-scavenging capacities of 74 compounds were used in the dataset. The SOAC values of compounds were obtained from previous studies [8]. The SOAC values can be calculated as the ratio of the quenching rate constants of the antioxidants and  $\alpha$ -tocopherol.

SOAC values for the compounds whose values are not available in literature but whose singlet-oxygen-scavenging rates are known were determined by calculating the ratio of the quenching rate constants of the antioxidant and  $\alpha$ -tocopherol. The natural logarithm of the rate ratio was used as the objective variable. Examples of compounds in the dataset are displayed in Figure 1. The names of all compounds in the dataset and their singlet-oxygen-scavenging capacities are listed in Table S1 of the Supporting Information.

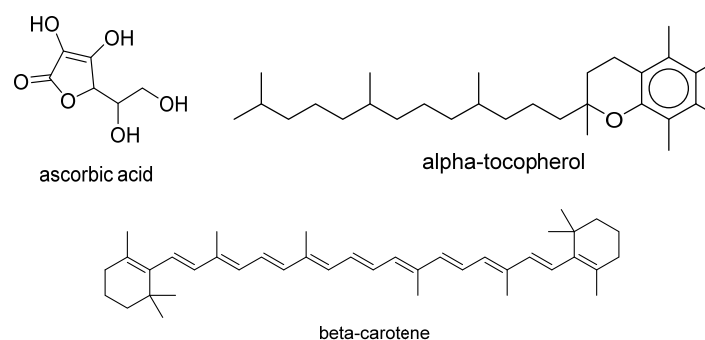


Figure 1. Examples of compounds in dataset.

The compounds data were obtained from PubChem as isomeric SMILES (Simplified Molecular Input Line Entry System) [18]. The most stable structure was determined as the initial structure for each sample by exploring the coordination from the molecular force field calculation using Balloon [19]. Then, the most stable structure was used as the initial structure. From the obtained structures, molecular descriptors and Morgan fingerprints were output using RDKit [20] and used as explanatory variables for the dataset. Using the most stable structure, PM7 [21] calculations were performed using MOPAC (Molecular Orbital PACkage) [21]. The PM7 method, which is a semiempirical method, was applied to optimize the structure under the assumption of vacuum conditions. We obtained HOMO (Highest Occupied Molecular Orbital), LUMO (Lowest Unoccupied Molecular Orbital), formation heat, and dipole moment values for each compound via PM7. Because PM6 is as accurate as density functional theory (DFT) calculations [22,23], we assumed that

the PM7 calculations exhibit identical acceptable errors as DFT calculations under the B3LYP/6-31G\* condition.

Feature selection was performed using more than 100 descriptors generated through RDkit and PM7. Firstly, descriptors with zero variance or duplicate content were deleted. Secondly, the correlation coefficients between explanatory variables were calculated. For the two descriptors with the highest absolute values of correlation coefficients, one of the descriptors was deleted. Descriptors with high correlation coefficients with many variables were omitted from the dataset because one indicator could substitute for many descriptors. Therefore, the number of descriptors used in the dataset was 61. The list of descriptors is provided in the Supporting Information.

## 2.2. Machine Learning Model

In this study, XGBoost [24,25], LightGBM [12,26], CatBoost [27], random forest [28], AdaBoost [29], LASSO regression [30], and the deep neural network (DNN) [31] were used to construct machine learning models. Additionally, we used Scikit-learn [32], TensorFlow [33], and Keras [34] to construct the machine learning model. The hyperparameters were set as described in the Supporting Information. The compound data were randomly categorized into 66 training data and 8 test data by setting the random state values to 0, 10, and 100. The same machine learning model was trained on molecular descriptors and Morgan fingerprints separately to evaluate and compare prediction performance. In this study, we compared the prediction accuracy using the coefficient of determination  $R^2$  and root mean squared error (RMSE) and used the leave-one-out cross-validation (LOOCV) method to ensure sufficient level of learning. In LOOCV, one sample from the original set was removed and used as the validation sample, and the rest was used for training. By changing a validation sample,  $n$  training datasets were created when the original training dataset had  $n$  samples. The output value by LOOCV ( $RMSE_{LOO}$ ) is the average value of the RMSE obtained by evaluating the machine learning model with each validation sample.

## 2.3. Feature Importance

The contribution of each feature to the prediction was analyzed by using the absolute values of coefficients obtained by LASSO regression and feature importance obtained by random forest, XGBoost, LightGBM, CatBoost, and AdaBoost. To analyze feature importance, we created a feature ranking for each machine learning model [16]. We focused on prediction models with  $R^2 \geq 0$  and analyzed the feature importance. For each model, the top-10 important features were assigned a score in the order of importance: 10, 9, 8, 7, 6, 5, 4, 3, 2, and 1. By summing the scores of the prediction models and comparing them, we identified the important variables in the machine learning model for antioxidant capacity.

## 3. Results

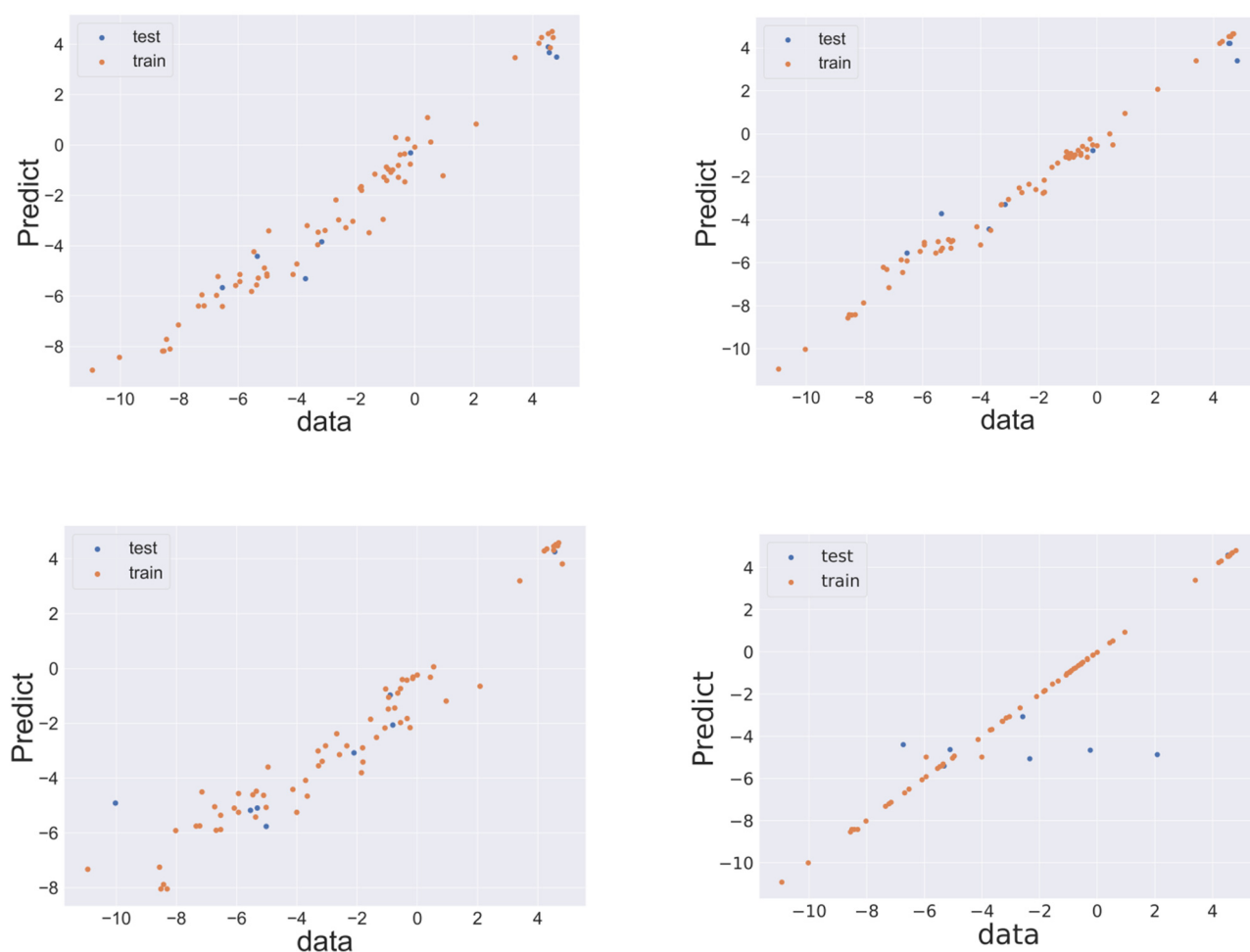
### 3.1. Prediction

The performance of machine learning was estimated using LOOCV. To evaluate the prediction performance of each algorithm,  $RMSE_{LOO}$  was calculated. Table 1 lists the values of the evaluation indices for prediction accuracy when predicting the test data for each model. Comparison of the datasets revealed that superior predictive performance was achieved when molecular descriptors were used as explanatory variables. In AdaBoost and LASSO regression, superior performance was obtained when Morgan fingerprints were used as the explanatory variable. The optimal performance was obtained when random forest was combined with molecular descriptors as explanatory variables. Random forest exhibited robustness of the prediction model as  $R^2$  did not become negative. When the Morgan fingerprint was used as an explanatory variable, the performance of AdaBoost was the highest, followed by random forest and XGBoost. In the case of AdaBoost, the coefficient of determination was sometimes negative, whereas  $R^2$  was greater than zero in the case of XGBoost or random forest. XGBoost and random forest exhibited prediction accuracy comparable to AdaBoost and robustness superior to that of AdaBoost. Figure 2

displays the relationship between the values of the target variables (literature values) and the predicted values of the datasets for the random forest model trained with molecular descriptors and the AdaBoost, random forest, and XGBoost models trained by Morgan fingerprints. The relationship between literature values and predicted values in other cases is shown in the Supporting Information.

**Table 1.** Regression metrics values of machine learning using estimator candidates.

Model	Dataset	Random_State	RMSE	R <sup>2</sup>	RMSE <sub>LOO</sub>
XGBoost	descriptors	0	1.2913	0.9147	1.7566
		10	1.6867	0.8279	
		100	2.0362	0.6968	
	fingerprint	0	1.0857	0.9397	1.7857
		10	2.7150	0.5542	
		100	3.1873	0.2325	
LightGBM	descriptors	0	2.4106	0.7028	1.8723
		10	1.8021	0.7546	
		100	1.8021	0.7546	
	fingerprint	0	3.7413	0.2841	2.4055
		10	2.9596	0.4702	
		100	3.4717	0.0894	
CatBoost	descriptors	0	2.0570	0.7836	1.7375
		10	2.1533	0.7196	
		100	2.4917	0.5309	
	fingerprint	0	2.1026	0.7739	2.5477
		10	2.3711	0.6600	
		100	3.6750	−0.0200	
Randomforest	descriptors	0	1.0123	0.9476	1.5731
		10	1.9342	0.7737	
		100	1.9166	0.7225	
	fingerprint	0	1.3950	0.9005	1.7613
		10	1.2796	0.9163	
		100	2.0136	0.7548	
AdaBoost	descriptors	0	0.8141	0.9661	1.7006
		10	2.1277	0.7262	
		100	1.9945	0.6995	
	fingerprint	0	0.9573	0.9531	1.6017
		10	1.8246	0.7986	
		100	3.8793	−0.1370	
LASSO	descriptors	0	2.1668	0.7599	2.3314
		10	2.4023	0.6510	
		100	5.5896	−1.3605	
	fingerprint	0	1.7224	0.8483	1.9853
		10	3.0045	0.4540	
		100	4.0502	−0.2393	
DNN	descriptors	0	2.0145	0.7924	2.9865
		10	2.6807	0.5653	
		100	3.6878	−0.0275	
	fingerprint	0	1.8309	0.8286	3.3584
		10	3.3159	0.3350	
		100	3.4791	0.0855	



**Figure 2.** Relationship between objective feature values in the dataset and the predicted values. **Top row:** descriptors/random forest (**left**), and Morgan fingerprint/AdaBoost (**right**); **Bottom row:** Morgan fingerprint/random forest (**left**), and Morgan fingerprint/XGBoost (**right**). In all panels, the blue and orange markers indicate test and training data, respectively.

### 3.2. Importance Analysis

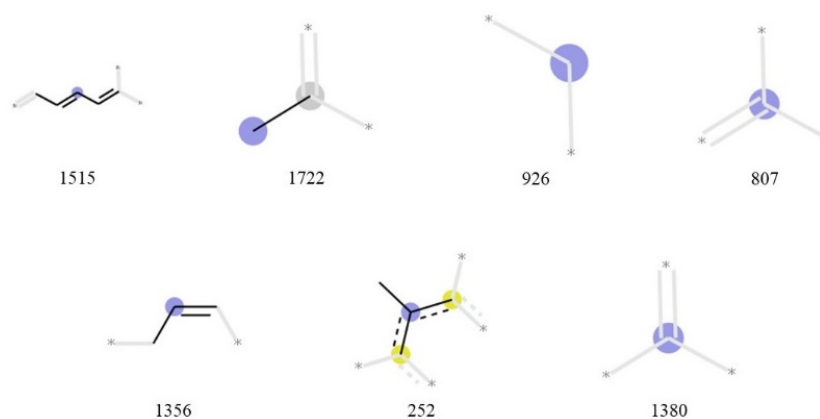
We focused on the sum of the importance scores of all models to understand the relationship between the structure and singlet-oxygen-scavenging capacity. The top seven features and their scores in the order of summation are listed in Table 2. The ranking of the importance obtained from each model is displayed in the Supporting Information. From the molecular descriptor dataset, HOMO, HOMO–LUMO gap, SlogP\_VSA2, SlogP\_VSA4, SlogP\_VSA6, and PEOE\_VSA7 values were obtained. SlogP\_VSA2, SlogP\_VSA4, SlogP\_VSA6 and PEOE\_VSA7 are molecular descriptors representing atomic contributions to logP and atomic partial charge [35]. They are based on van der Waals surface area. BalabanJ is a topological index based on the sum of distance of each bond in a certain molecule [36].

Figure 3 displays the partial structure of each bit considered important in the machine learning using the Morgan fingerprint. Bits 1515 and 252 indicate that the presence or absence of a conjugated system considerably affects the singlet-oxygen-scavenging ability. Bit 1722 represents a methyl group attached to an aliphatic carbon atom, and bit 926 represents a carbon chain or a carbon atom with two C–H bonds. Bits 807, 1356, and 1380 represent carbon atoms with double bonds and single bonds. Bits 807 and 1380 exhibit no information regarding bond destination of each carbon atom.

**Table 2.** Feature importance points of features with a large contribution for prediction.

Descriptor	Score	Fingerprint	Score
HOMO	175.5	1515	148.5
HOMO–LUMO gap	109	1722	118.5
SlogP_VSA4	84.5	926	93.5
SlogP_VSA2	72.5	807	87.5
SlogP_VSA6	72	1356	80
PEOE_VSA7	62	252	58.5
BalabanJ	44	1380	35.5

HOMO, Highest Occupied Molecular Orbital; LUMO, Lowest Unoccupied Molecular Orbital; VSA, Van der Waals Surface Area; PEOE, Partial Equation of Orbital Electronegativity.



**Figure 3.** Important Morgan fingerprint bits. Bit 1515 suggests conjugate systems in a molecule. Bit 1722 represents a methyl group. Bits 926, 807, and 1380 represent carbon atoms. Bit 1356 suggests an allyl group or an unsaturated aliphatic carbon ring. Bit 252 represents an aromatic ring that has a methyl group.

## 4. Discussion

### 4.1. Prediction Accuracy

An excellent prediction performance was obtained by using the proposed machine learning method. Predictions without computational chemistry and using computational chemistry were both possible, which indicated that machine learning can be useful for understanding the antioxidant capacity. Machine learning exhibits considerable flexibility to provide prediction accuracy according to the user's objectives by using various explanatory variables, which is critical for practical applications. Each bit of the Morgan fingerprint is represented by 0 and 1, respectively. Furthermore, the substructure from the central atom to at most two bonds ahead is used as the feature value. Therefore, the length of the conjugated system is not reflected, and counting the number of locations that have the same substructure is not feasible. Therefore, the prediction results using the Morgan fingerprint were less accurate than those using molecular descriptors. LASSO regression was an exception to this trend, but more variables can improve prediction accuracy. DNNs did not perform well on either dataset in this study. Thus, the simple DNN was not suitable for predicting the antioxidant capacity using a small amount of data.

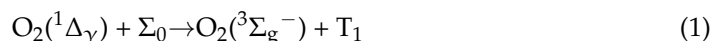
Comparing the RMSE during cross-validation with the RMSE when evaluated on the test data, we determined that the machine learning models did not overfit. We also revealed that the Morgan fingerprint tends to overfit our models. Whether or not overfitting occurs depends on the size of the data, especially the number of features.

### 4.2. Importance Analysis

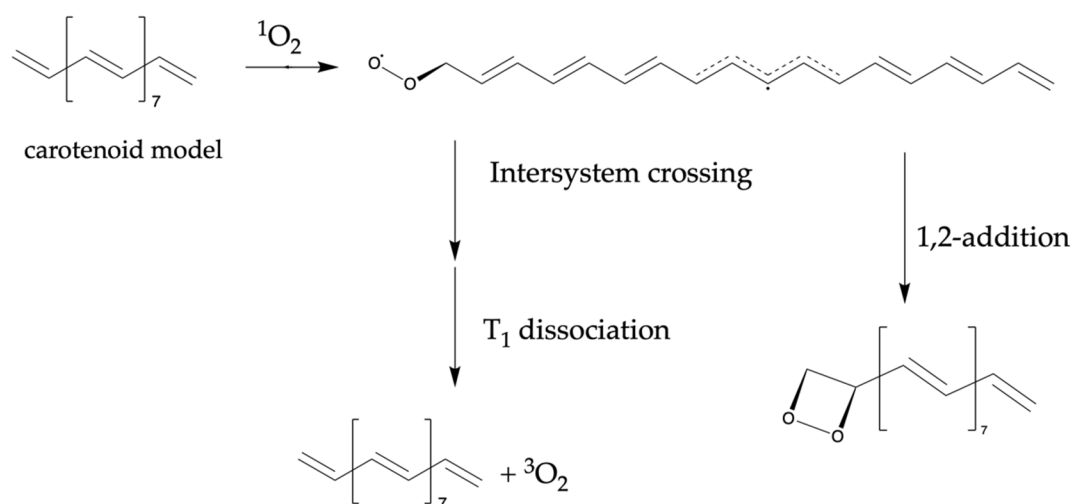
The results of the Morgan fingerprint revealed that conjugated systems and carbon atoms with double and single bonds are of particular importance. As the conjugated chain became longer, the absorption due to the  $\pi$ - $\pi^*$  transition shifted to the longer wavelength

side [37]. This result indicated that the HOMO–LUMO gap became narrower, which is consistent with the high importance of the HOMO and HOMO–LUMO gaps in molecular descriptor datasets.

The electronic energy transfer (EET) [17] is expressed as Equation (1), in which the quencher transitions to the triplet state upon elimination of singlet oxygen and exhibits a rate constant close to the diffusion-rate-limiting rate of the quenching mechanism of singlet oxygen.



In this scheme, the narrow HOMO–LUMO gap implied the ease of energy exchange between the quencher and singlet oxygen. In the EET mechanism, an encounter complex is formed by singlet oxygen and an antioxidant in the singlet state, and energy transfer is proposed to occur through the term crossing of the complex [38]. The larger HOMO value suggested that the antioxidant is more likely to approach the singlet oxygen, which is an electrophilic agent, thus promoting energy transfer. The reaction mechanism reported as a competitive reaction in this scheme is displayed in Scheme 1 [39]. The quencher and oxygen reacted to form a complex that underwent radicalization. Subsequently, oxygen chemical quenching occurred, or peroxides and carbonyl compounds were formed. Therefore, HOMO is expected to be used as an indicator of nucleophilicity from antioxidants to oxygen during chemical quenching.



**Scheme 1.** Competitive reactions in electron energy transfer (EET) mechanisms considered as chemical quenching.

In the quenching of singlet oxygen by phenols, two types of physical quenching reactions are known, namely, electrons are transferred between the aromatic ring and oxygen in the transition state but no oxygen is consumed, and chemical quenching, in which peroxides are formed [40]. The EET mechanism was consistent with machine learning inference because the reaction rate of the EET mechanism is close to the diffusion rate.

As mentioned earlier, singlet-oxygen-scavenging activity is correlated with the length of the conjugated chain, the length of the carbon chain, and the absorption wavelength of the ground state. The Morgan fingerprint bit displayed in Figure 3 is a critical indicator that can be explained using the EET mechanism as well as the HOMO and HOMO–LUMO gap because it can represent the structure of the conjugated system or a part of it. Because fingerprints can be used for machine learning, the substructure of the compound could be used as an alternative indicator to the HOMO.

Because SlogP\_VSA2, SlogP\_VSA4, SlogP\_VSA6, and PEOE\_VSA7 are critical when molecular descriptors are used as a dataset, atomic distribution for solubility, and partial charge were important in predicting antioxidant capacity. Reorganizing the dataset and examining new descriptors of electron density or polarity can improve the prediction performance and reveal electronic effects that are critical for studying antioxidants.

Analyzing the behavior of machine learning models by feature importance can explain prediction accuracy. Although feature importance is an ineffective measure for explaining causality, we interpreted it chemically by comparing it to previously known information. The process of testing hypotheses formulated by machine learning with computational chemistry and experiments is useful not only for efficiently evaluating properties that previously relied solely on experiments, such as antioxidant capacity, but also for verifying the validity of the evaluation.

## 5. Conclusions

A critical challenge in applying machine learning in chemistry and life science is that the prediction process remains unclear and the amount of data to be collected is small. In this study, we developed a prediction model that is easily interpretable by chemists and requires only a small amount of data. The proposed machine learning model can predict singlet-oxygen-scavenging activity of compounds, which is critical in food science. Molecular descriptors and Morgan fingerprints were used to validate the simple antioxidant capacity by the proposed method, and the importance of the features and interpreted the behavior of the machine learning model were examined chemically. Thus, the evaluation of antioxidant capacity was simplified and did not require time-consuming experiments. The prediction mechanism was also explained.

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/article/10.3390/antiox10111751/s1>, Table S1: The Dataset; Figures S1–S5: Relationship between the objective variable of the data set and its predictive values; Table S2: Feature importance in ensemble learning and coefficient of linear regression.

**Author Contributions:** All authors contributed to the project, and the main contributions are as follows: conceptualization, T.F. and H.G.; methodology, T.F.; validation, T.F.; investigation, T.F.; resources, H.G.; writing—original draft preparation, T.F.; writing—review and editing, H.G.; visualization, T.F.; supervision, H.G.; project administration, H.G.; funding acquisition, H.G. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** All of the data is contained within the article and the Supplementary Materials.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Lee, J.; Koo, N.; Min, D.B. Reactive Oxygen Species, Aging, and Antioxidative Nutraceuticals. *Compr. Rev. Food Sci. Food Saf.* **2004**, *3*, 21–33. [[CrossRef](#)] [[PubMed](#)]
2. Kaur, C.; Kapoor, H.C. Antioxidants in Fruits and Vegetables—The Millennium’s Health. *Int. J. Food Sci. Technol.* **2001**, *36*, 703–725. [[CrossRef](#)]
3. Cao, G.; Alessio, H.M.; Cutler, R.G. Oxygen-Radical Absorbance Capacity Assay for Antioxidants. *Free Radic. Biol. Med.* **1993**, *14*, 303–311. [[CrossRef](#)]
4. Ou, B.; Hampsch-Woodill, M.; Prior, R.L. Development and Validation of an Improved Oxygen Radical Absorbance Capacity Assay Using Fluorescein as the Fluorescent Probe. *J. Agric. Food Chem.* **2001**, *49*, 4619–4626. [[CrossRef](#)]
5. Rey, F.; Zacarias, L.; Rodrigo, M.J. Carotenoids, Vitamin C, and Antioxidant Capacity in the Peel of Mandarin Fruit in Relation to the Susceptibility to Chilling Injury during Postharvest Cold Storage. *Antioxidants* **2020**, *9*, 1296. [[CrossRef](#)]
6. Edge, R.; McGarvey, D.J.; Truscott, T.G. The Carotenoids as Anti-Oxidants—A Review. *J. Photochem. Photobiol. B* **1997**, *41*, 189–200. [[CrossRef](#)]
7. Takahashi, S.; Tsutsumi, A.; Aizawa, K.; Suganuma, H. Daily Radical Scavenging and Singlet Oxygen Quenching Capacity Intake from Fruits and Vegetables in Japan. *Food Sci. Technol. Res.* **2018**, *24*, 921–933. [[CrossRef](#)]
8. Mukai, K. Antioxidant Activity of Foods: Development of Singlet Oxygen Absorption Capacity (SOAC) Assay Method. *J. Nutr. Sci. Vitaminol.* **2019**, *65*, 285–302. [[CrossRef](#)]
9. Antioxidant-Function. Available online: <http://www.antioxidant-function.com/detaile/> (accessed on 1 November 2021).



10. Jordan, M.I.; Mitchell, T.M. Machine Learning: Trends, Perspectives, and Prospects. *Science* **2015**, *349*, 255–260. [[CrossRef](#)]
11. Lutnick, B.; Ginley, B.; Govind, D.; McGarry, S.D.; LaViolette, P.S.; Yacoub, R.; Jain, S.; Tomaszewski, J.E.; Jen, K.Y.; Sarder, P. An Integrated Iterative Annotation Technique for Easing Neural Network Training in Medical Image Analysis. *Nat. Mach. Intell.* **2019**, *1*, 112–119. [[CrossRef](#)]
12. Zhang, J.; Mucs, D.; Norinder, U.; Svensson, F. LightGBM: An Effective and Scalable Algorithm for Prediction of Chemical Toxicity—Application to the Tox21 and Mutagenicity Data Sets. *J. Chem. Inf. Model.* **2019**, *59*, 4150–4158. [[CrossRef](#)] [[PubMed](#)]
13. Liu, B.; Ramsundar, B.; Kawthekar, P.; Shi, J.; Gomes, J.; Luu Nguyen, Q.; Ho, S.; Sloane, J.; Wender, P.; Pande, V. Retrosynthetic Reaction Prediction Using Neural Sequence-to-Sequence Models. *ACS Cent. Sci.* **2017**, *3*, 1103–1113. [[CrossRef](#)]
14. Podio, N.S.; López-Froilán, R.; Ramirez-Moreno, E.; Bertrand, L.; Baroni, M.V.; Pérez-Rodríguez, M.L.; Sánchez-Mata, M.C.; Wunderlin, D.A. Matching In Vitro Bioaccessibility of Polyphenols and Antioxidant Capacity of Soluble Coffee by Boosted Regression Trees. *J. Agric. Food Chem.* **2015**, *63*, 9572–9582. [[CrossRef](#)] [[PubMed](#)]
15. Rogers, D.; Hahn, M. Extended-Connectivity Fingerprints. *J. Chem. Inf. Model.* **2010**, *50*, 742–754. [[CrossRef](#)]
16. Chen, C.H.; Tanaka, K.; Kotera, M.; Funatsu, K. Comparison and Improvement of the Predictability and Interpretability with Ensemble Learning Models in QSPR Applications. *J. Cheminform.* **2020**, *12*, 19. [[CrossRef](#)]
17. Schweitzer, C.; Schmidt, R. Physical Mechanisms of Generation and Deactivation of Singlet Oxygen. *Chem. Rev.* **2003**, *103*, 1685–1757. [[CrossRef](#)] [[PubMed](#)]
18. Weininger, D. SMILES, a Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules. *J. Chem. Inf. Model.* **1988**, *28*, 31–36. [[CrossRef](#)]
19. Vainio, M.J.; Johnson, M.S. Generating Conformer Ensembles Using a Multiobjective Genetic Algorithm. *J. Chem. Inf. Model.* **2007**, *47*, 2462–2474. [[CrossRef](#)] [[PubMed](#)]
20. RDKit. Available online: <https://www.rdkit.org/> (accessed on 1 November 2021).
21. Stewart, J.J.P. Optimization of Parameters for Semiempirical Methods VI: More Modifications to the NDDO Approximations and re-Optimization of Parameters. *J. Mol. Model.* **2013**, *19*, 1–32. [[CrossRef](#)]
22. Amić, D.; Lučić, B. Reliability of Bond Dissociation Enthalpy Calculated by the PM6 Method and Experimental TEAC Values in Antiradical QSAR of Flavonoids. *Bioorg. Med. Chem.* **2010**, *18*, 28–35. [[CrossRef](#)] [[PubMed](#)]
23. Nakata, M.; Shimazaki, T.; Hashimoto, M.; Maeda, T. PubChemQC PM6: Data Sets of 221 Million Molecules with Optimized Molecular Geometries and Electronic Properties. *J. Chem. Inf. Model.* **2020**, *60*, 5891–5899. [[CrossRef](#)]
24. Chen, T.; Guestrin, C. XGBoost. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
25. Sheridan, R.P.; Wang, W.M.; Liaw, A.; Ma, J.; Gifford, E.M. Extreme Gradient Boosting as a Method for Quantitative Structure–Activity Relationships. *J. Chem. Inf. Model.* **2016**, *56*, 2353–2360. [[CrossRef](#)]
26. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.Y. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 3147–3157.
27. Prokhorenkova, L.; Gusev, G.; Vorobev, A.; Dorogush, A.V.; Gulin, A. CatBoost: Unbiased Boosting with Categorical Features. *Adv. Neural Inf. Process. Syst.* **2018**, *31*, 6638–6648.
28. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
29. Freund, Y.; Schapire, R.E. A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *J. Comput. Syst. Sci.* **1997**, *55*, 119–139. [[CrossRef](#)]
30. Friedman, J.; Hastie, T.; Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J. Stat. Softw.* **2010**, *33*, 1–22. [[CrossRef](#)]
31. Ma, J.; Sheridan, R.P.; Liaw, A.; Dahl, G.E.; Svetnik, V. Deep Neural Nets as a Method for Quantitative Structure–Activity Relationships. *J. Chem. Inf. Model.* **2015**, *55*, 263–274. [[CrossRef](#)]
32. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
33. Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G.; Davis, A.; Dean, J.; Devin, M.; et al. Tensor flow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. *arXiv* **2015**, arXiv:1603.04467.
34. Gillet, F. Keras. 2015. Available online: <https://keras.io/> (accessed on 1 November 2021).
35. Labute, P. A Widely Applicable Set of Descriptors. *J. Mol. Graph. Model.* **2000**, *18*, 464–477. [[CrossRef](#)]
36. Balaban, A.T.; Balaban, A.T. Highly discriminating distance-based topological index. *Chem. Phys. Lett.* **1982**, *89*, 399–404. [[CrossRef](#)]
37. Edge, R.; Truscott, T.G. Singlet Oxygen and Free Radical Reactions of Retinoids and Carotenoids—A Review. *Antioxidants* **2018**, *7*, 5. [[CrossRef](#)] [[PubMed](#)]
38. Schmidt, R. Deactivation of O<sub>2</sub>(<sup>1</sup>Δg) Singlet Oxygen by Carotenoids: Internal Conversion of Excited Encounter Complexes. *J. Phys. Chem. A* **2004**, *108*, 5509–5513. [[CrossRef](#)]
39. Garavelli, M.; Bernardi, F.; Olivucci, M.; Robb, M.A. DFT Study of the Reactions between Singlet-Oxygen and a Carotenoid Model. *J. Am. Chem. Soc.* **1998**, *120*, 10210–10222. [[CrossRef](#)]
40. Al-Nu'airat, J.; Dlugogorski, B.Z.; Gao, X.; Zeinali, N.; Skut, J.; Westmoreland, P.R.; Oluwoye, I.; Altarawneh, M. Reaction of Phenol with Singlet Oxygen. *Phys. Chem. Chem. Phys.* **2018**, *21*, 171–183. [[CrossRef](#)]