*toxics*

MDPI

*Article*

# Predicting Aqueous-Phase Hydroxyl Radical Reaction Rate Constants Using Machine Learning

**Tongtong Shou [1], Ziyan Zou [1], Jingyi Yang [2], Ning Ruan [2], Yunjiang Zhang [1], Yun Wu [1] and Xinlei Ge [2,*]**

[1] Jiangsu Key Laboratory of Intelligent Atmospheric Environment Monitoring and Carbon–Pollution Co-Control, School of Environmental Science and Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China
[2] School of Energy and Environment, Southeast University, Nanjing 211189, China
* Correspondence: xinlei@seu.edu.cn

**Abstract**

Aqueous-phase hydroxyl radical (•OH) reaction rate constants control the environmental fate of organic compounds, yet experimental measurements remain unavailable for many compounds. Using 1229 organic species, this work established a machine learning framework that evaluates nine combinations of three molecular representations and three algorithms. All model selections used only five-fold cross-validation on the training set and out-of-fold (OOF) predictions; a held-out dataset was reserved for final evaluation. The molecular access system with random forest (MACCS–RF) combined model showed the best overall performance on the held-out test set with a coefficient of determination ($R^2$) of 0.7001, a root mean square error (RMSE) of 0.0815, and a mean absolute error of 0.0567. OOF analysis identified several structural classes with relatively large prediction errors. The local model for sulfur-containing compounds slightly improved the overall results, increasing $R^2$ to 0.7115 and reducing RMSE to 0.0799. SHapley Additive exPlanations (SHAP) analysis revealed that aromatic, cyclic, C=C-related, and sulfur-containing structural motifs contribute to •OH reactivity, while halogen and O–C–O substructures yielded negative contributions. Assessments of application domain (AD) using Tanimoto similarity and Williams plots indicated MACCS keys delivered much wider reliable coverage than Morgan fingerprints at the same thresholds. This study reveals structure-specific prediction biases for aqueous •OH reactivity and shows local modeling provides limited subclass-dependent benefits. By comparing chemical space coverage of different molecular representations, we propose an integrated workflow combining error diagnosis, structural interpretation, and AD evaluation for environmental kinetic quantitative structure–activity relationship (QSAR) modeling.

**Keywords:** hydroxyl radicals; aqueous phase reaction kinetics; machine learning; molecular fingerprints; applicability domain

Check for updates

Academic Editor: Mingce Long

Received: 11 October 2026
Revised: 12 November 2026
Accepted: 16 November 2026
Published: 18 November 2026

**Copyright:** © 2026 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.

## 1. Introduction

The hydroxyl radical (•OH) is a highly reactive oxidizing species that plays an important role in the transformation of organic compounds in natural waters, engineered water treatment systems, and atmospheric droplets. The aqueous-phase •OH reaction rate constant ($k_{OH}$) is therefore a fundamental kinetic parameter for evaluating reactivity, persistence, and behavior of the organic compounds. Extensive kinetic compilations and mechanistic studies have established the central role of •OH reactions in aqueous-phase chemistry [1,2]. However, experimental determination of these rate constants requires

specialized kinetic techniques, and measurements are still unavailable for a large number of environmentally relevant organic compounds.

To address gaps in experimental data, quantitative structure-property/activity relationship (QSPR/QSAR) models have been developed to predict aqueous •OH reaction rate constants based on molecular structures [3]. Earlier studies used molecular descriptors and conventional statistical or nonlinear regressions to predict $k_{OH}$ for diverse organic compounds [4–6]. Recent modern QSPR frameworks, on the other hand, increasingly integrated machine learning (ML) methods to improve prediction performance [7,8]. For instance, Zhong et al. developed explainable ML and molecular image models to predict aqueous •OH reactivity [9,10]. Sanches-Neto et al. developed a framework integrating molecular fingerprints, ML algorithms, external validation, application domain (AD) assessment, and SHapley Additive exPlanations (SHAP) interpretation [11]. Recent studies have also explored multi-task and semi-supervised learning to expand the predictive coverage of aqueous-phase radical kinetic parameters [12,13]. Previous reviews have summarized the use of QSPR approaches to predict the reactivity of organic compounds with radicals in aqueous phases [14]. Moreover, Jin et al. [15] provided experimental kinetic data for organic micropollutants, and they further developed a QSPR model using molecular descriptors and evaluated its predictive performance and AD [16]. Mao et al. [17] introduced a deep-learning framework combining different molecular representations for aqueous •OH reaction rate prediction. Liang et al. [18] applied ML, SHAP interpretation, and AD assessment to aqueous halogen radical reactions. More recently, Huang et al. [19] explored knowledge transfer across different radical reaction systems using graph neural networks. Together, these studies demonstrate progress in kinetic data collection, molecular representation, ML prediction, and model reliability assessment.

The Organization for Economic Co-operation and Development (OECD) principles for QSAR validation, originally for regulatory chemical assessment, require a defined endpoint, an unambiguous algorithm, a defined AD, and appropriate measures of goodness-of-fit, robustness, and predictivity [20]. More broadly, rigorous internal and external validation, careful separation of model selection from final performance estimation, and explicit characterization of the AD are considered essential for reliable QSPR/QSAR modeling [21–23]. These requirements are particularly relevant to aqueous-phase $k_{OH}$ prediction because a model with acceptable overall performance may still show systematic biases in specific regions of chemical space or for particular structural classes. Identifying such high-error regions and evaluating class-specific models under a training-only selection framework may therefore offer a practical strategy to improve model reliability. While prior modeling studies have attained reasonable predictive performance, differences in prediction errors across structural classes remain under-explored [15–19]. It is unclear whether compounds from high-error groups can gain benefits from targeted local models. Resolving these points can reveal the limitations of global models and enhance prediction reliability for specific sets of organics.

In this regard, this study established a structurally diverse dataset of aqueous-phase •OH reaction kinetics to compare the performance of models combining three molecular representations (MRs) with three ML algorithms. This framework is designed to prioritize model reliability by integrating predictive performance evaluation, heterogeneity in structural-class error, targeted local optimization, SHAP interpretation, and chemical-space characterization of AD.

## 2. Materials and Methods

### *2.1. Dataset and Characteristics*

We compiled aqueous-phase reaction rate constants for organic compounds reacting with hydroxyl radicals (•OH) at room temperature. Temperature, pH, and molecular protonation state can all influence aqueous-phase •OH kinetics, especially for ionizable species [24,25]. The kinetic data were retrieved from peer-reviewed experimental studies, relevant kinetic review articles, and the National Institute of Standards and Technology (NIST) database [11,25–30]. In total, 1726 reaction rate constants were initially collected. The Simplified Molecular-Input Line-Entry System (SMILES) strings for each compound were obtained using their Chemical Abstracts Service (CAS) registry numbers, and RDKit (version 2023.09.6; RDKit open-source cheminformatics project, https://www.rdkit.org) was adopted for molecular structure parsing and standardization. After screening based on experimental and structural information, 1659 records were retained, of which 1658 yielded valid molecular structures. The International Chemical Identifier (InChI) was subsequently used to detect duplicated molecular structures and flag redundant entries. When multiple room temperature rate constants were reported for the same compound, their arithmetic mean was adopted as the final target value. After such data preprocessing, 1229 organics were retained for ML modeling. The data curation procedures and the number of records retained at each step are summarized in Table S1 of the Supplementary Materials. The rate constants were then natural-log transformed and normalized to the scale of 0–1 to narrow the dynamic range and improve the symmetry of the response distribution.

In addition, to better characterize the chemical space of the dataset, RDKit was used to calculate a series of fundamental molecular properties, including molecular weight (MW), octanol–water partition coefficient (MolLogP), Topological Polar Surface Area (TPSA), number of Hydrogen-Bond Donors (HBD) and Hydrogen-Bond Acceptors (HBA), number of rotatable bonds, total ring count, aromatic ring count, and fraction of $sp^3$-hybridized carbon atoms ($FractionC_{SP3}$). The compounds were subsequently classified on the basis of their molecular scaffolds and major functional groups to depict dataset composition and assess variations in model performance across different structural classes.

### *2.2. Molecular Representations (MRs) and Feature Construction*

In this study, the molecular structures were represented using Morgan fingerprints, Molecular ACCess System (MACCS) keys, and RDKit 2D descriptors. Morgan extended connectivity fingerprints extract local substructural information by probing the neighboring bonding environment of atoms and are widely used for molecular-property modeling [31,32]. MACCS keys capture predefined structural motifs and functional-group information [33]. Molecular structure standardization and descriptor calculation were implemented via RDKit-based cheminformatics workflows [34,35]. To evaluate the effect of Morgan fingerprint length on model performance, fingerprints with 512, 1024, 2048, 3072, 4096, and 8192 bits were generated. Fingerprint-length tuning was performed exclusively within the training set using five-fold cross-validation (CV). To avoid data leakage, the held-out test set was not used to determine the optimal fingerprint length. Compared with Morgan fingerprints, MACCS keys have lower dimensionality and better structural interpretability, making them suitable for both prediction and subsequent structural interpretation. RDKit 2D descriptors provide continuous numerical representations of molecular composition, topology, and physicochemical characteristics. Descriptors suffering from calculation failures, invalid entries, or poor information were removed during preprocessing, and the refined feature sets were fed as inputs to the different ML models.

### *2.3. Machine Learning (ML) Algorithms and Evaluation of Performance*

Three ML algorithms, including Random Forest (RF), eXtreme Gradient Boosting (XGBoost), and Support Vector Regression (SVR), were adopted to model the quantitative relationship between molecular structures and aqueous •OH reaction rate constants. The three algorithms are widely used nonlinear learning approaches [36–38], and the entire modeling workflow was implemented using Python (version 3.9; Python Software Foundation, Wilmington, DE, USA), with scikit-learn (version 1.6.1) and XGBoost (version 2.1.4) [39]. The 1229 compounds were randomly divided into a training set ($n = 983$) and a held-out test set ($n = 246$) at an 8:2 ratio using a fixed random seed of 521. All model-related decisions, including fingerprint-length tuning, hyperparameter optimization, global-model selection, structural-class diagnosis, class-specific model selection, hierarchical routing-rule development, and determination of application domain (AD) threshold, were restricted only to the training set via five-fold CV and out-of-fold (OOF) predictions (details in the following sections). The held-out test set was completely withheld and only used for final performance assessment. Once global, local models and routing strategy were finalized from training-set data, models were refitted on the full training dataset prior to final evaluation.

During the CV, for each fold, 4/5 of the training compounds were used for model fitting and the remaining 1/5 data for validation; accordingly, every compound in the training set generated an OOF prediction from a model trained without that compound. Cross-validated performance and OOF predictions were used for model comparison and subsequent structural-class analysis, strictly following the rule that separate model selection from final error estimation [22,40]. Combining three molecular representations with the three ML algorithms yielded nine candidate models. Model performances were evaluated using the coefficient of determination ($R^2$), root mean square error (RMSE), and mean absolute error (MAE), as defined in Equations (1), (2), and (3), respectively. Higher $R^2$ and lower RMSE and MAE values indicate better performance.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_{\mathrm{exp},i} - y_{\mathrm{pred},i}\right)^2}{\sum_{i=1}^{n}\left(y_{\mathrm{exp},i} - \bar{y}_{\mathrm{exp}}\right)^2} \quad (1)$$

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}\left(y_{\mathrm{exp},i} - y_{\mathrm{pred},i}\right)^2}{n}} \quad (2)$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|y_{\mathrm{exp},i} - y_{\mathrm{pred},i}\right| \quad (3)$$

where $y_{\mathrm{exp},i}$ and $y_{\mathrm{pred},i}$ represent the experimental and predicted values for the $i$-th compound, respectively; $\bar{y}_{\mathrm{exp}}$ represents the mean of the experimental values; and $n$ is the total number of compounds.

While the molecular representations and ML algorithms used here are well-established, their sophisticated integration constitutes a key methodological feature. Following the training-only protocols, we systematically evaluated nine models for more exhaustive configuration comparisons than most prior studies. All model decisions (hyperparameter tuning, structural-class identification, local-model routing) relied exclusively on training-set CV and OOF predictions; the held-out test set was used only for final evaluation and excluded from screening or optimization.

### *2.4. Prediction of Errors Across Structural Classes and Class-Specific Models*

Since a compound may belong to multiple classes, a multi-label structural classification scheme was adopted for our dataset. For example, a halogenated aromatic carboxylic acid can simultaneously be classified as aromatic, cyclic, halogen-containing, carbonyl-containing, and carboxylic acid. To detect structural domains where the global model yields

relatively large prediction errors, structural-class error analysis was conducted using OOF predictions derived from the training set. For each class, we calculated the number of compounds, RMSE, MAE, fraction of large-error compounds, and the ratio of class-specific RMSE to the overall OOF RMSE.

For each key high-error structural class identified from the training set OOF analysis, the nine candidate models were evaluated using only the corresponding training compounds. Hyperparameter optimization and local-model selection strictly followed the training-only protocol too. Five-fold CV was used for model comparison and hyperparameter optimization; cross-validated RMSE served as the primary model-selection criterion, with other error metrics used as supporting evidence. Next, a structure-guided hierarchical routing rule was established. A structural class was incorporated into the hierarchical strategy only if the selected class-specific model showed consistent improvement over the global model according to the predefined CV criteria derived from the training set. After completing all routing decisions, the global and class-specific models were refitted with the full training compounds.

The predefined hierarchical strategy was then applied without modification to the held-out test set. Compounds belonging to structural classes selected for local optimization were routed to the corresponding class-specific models, whereas all remaining compounds were predicted using the global model. Final held-out performance was evaluated using $R^2$, RMSE, MAE, and Pearson's correlation coefficient. In addition, 5000 paired bootstrap resamples of the held-out predictions were generated after the routing rule was determined to characterize the uncertainty and stability of the observed differences in RMSE and MAE; the bootstrap results did not affect model or routing selection [41].

### *2.5. Model Interpretation*

To elucidate the key structural features governing aqueous •OH reaction rate constants, SHAP analysis was used to interpret the final global model. SHAP provides additive feature-attribution values rooted in cooperative game theory and enables the characterization of both local and global model behavior [42–44]. Since MACCS-RF was selected as the global model in this study (See Section 3.2), SHAP analysis was performed on the MACCS-RF model refitted with the complete training set. The obtained SHAP values were used to assess the relative importance and contributing direction of individual MACCS structural keys.

### *2.6. Applicability Domain (AD)*

The AD analysis was used to determine whether target compounds are adequately represented within the chemical space covered by the training data and to flag predictions requiring cautious interpretation, following the established QSAR validation and AD principles [45,46]. For Morgan fingerprints and MACCS keys, structural similarity was quantified by the Tanimoto coefficient, a standard metric for binary molecular fingerprints [47]. For two fingerprints denoted $A$ and $B$, the Tanimoto coefficient is calculated as follows:

$$T_c(A, B) = \frac{c}{a + b - c} \tag{4}$$

where $a$ and $b$ signify the numbers of bits set to 1 in fingerprints $A$ and $B$, respectively, and $c$ represents the number of bits set to 1 in both fingerprints [47]. During the final held-out evaluation, the Tanimoto similarity of each test compound to all compounds in the full training set was calculated, and the value of maximum similarity ($S_i$) was adopted as the metric for that compound:

$$S_i = \max_{j \in \text{train}} T_c(i, j) \tag{5}$$

The candidate Tanimoto similarity thresholds were also evaluated only on the training data within the five-fold CV framework. For each fold, the maximum similarity of every subset to the corresponding four-fold training subset was calculated, and the coverage and RMSE of validation compounds satisfying each candidate threshold were summarized across folds. The similarity threshold was chosen from these training-set validation results by jointly considering coverage and prediction error. The threshold was then fixed and applied without modification to the held-out test set. A held-out test compound was regarded as falling within the fingerprint-based AD if its maximum similarity to the full training set was greater than or equal to the predefined threshold.

For RDKit 2D descriptors, the model's AD was assessed via the Williams plot. This approach determines whether a sample falls within the model's AD based primarily on its leverage value and standardized residual [45,46]. The leverage value $h_i$ for compound $i$ quantifies the extent to which the compound deviates from the descriptor space spanned by the training set and is defined as follows:

$$h_i = x_i^T \left(\mathbf{X}^T \mathbf{X}\right)^{-1} \mathbf{x_i} \quad (6)$$

$$h^* = \frac{3(p+1)}{n} \quad (7)$$

$$SR_i = \frac{e_i}{s\sqrt{1-h_i}} \quad (8)$$

where $x_i$ denotes the descriptor vector of compound $i$, $\boldsymbol{X}$ is the descriptor matrix of the training set, $p$ is the number of descriptors retained for model construction and AD analysis, $n$ is the number of training compounds, and $s$ is the standard error of the training-set residuals. A compound with $h_i > h^*$ was treated to occupy a high-leverage extrapolation regime, while $|SR_i| > 3$ identifies a potential response outlier.

## 3. Results and Discussion

### *3.1. Dataset Characteristics and Chemical Space Distribution*

Our final dataset contains 1229 organic compounds. The aqueous •OH rate constants ranged from $2.0 \times 10^6$ to $3.8 \times 10^{10}$ $M^{-1}$ $s^{-1}$, spanning approximately four orders of magnitude. The first quartile (Q1), median, and third quartile (Q3) were $1.7 \times 10^9$, $4.6 \times 10^9$, and $7.9 \times 10^9$ $M^{-1}$ $s^{-1}$, respectively, while the mean was $5.77 \times 10^9$ $M^{-1}$ $s^{-1}$. As shown in Figure 1, the original $k_{OH}$ value showed a pronounced skewed distribution. After natural logarithmic transformation, the distribution became more concentrated, with most $ln(k_{OH})$ values ranging from 21 to 23, although a long tail remained toward the low-reactivity region.

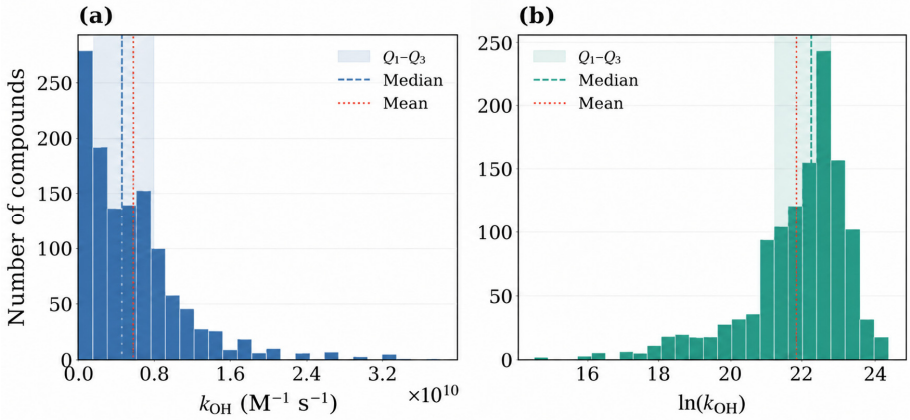



**Figure 1.** Distribution of aqueous-phase •OH reaction rate constants. (**a**) Original values ($k_{OH}$); (**b**) Natural log-transformed values ($ln(k_{OH})$).

Figure 2 shows the distributions of key physicochemical properties in the final dataset. MWs of compounds are mostly below 400. MolLogP spans hydrophilic to hydrophobic regimes with a peak near 2, while TPSA values are mostly below 200 Å$^2$. Distributions of HBD, HBA and rotatable bonds are dominated by low values (<10). Most compounds contain 0–2 total rings and 0 or 1 aromatic ring. Fraction$C_{SP3}$ shows an accumulated distribution near 0 and 1, reflecting plentiful unsaturated-aromatic and saturated-aliphatic structures. Overall, these properties reflect the large structural and physicochemical diversity of the dataset.

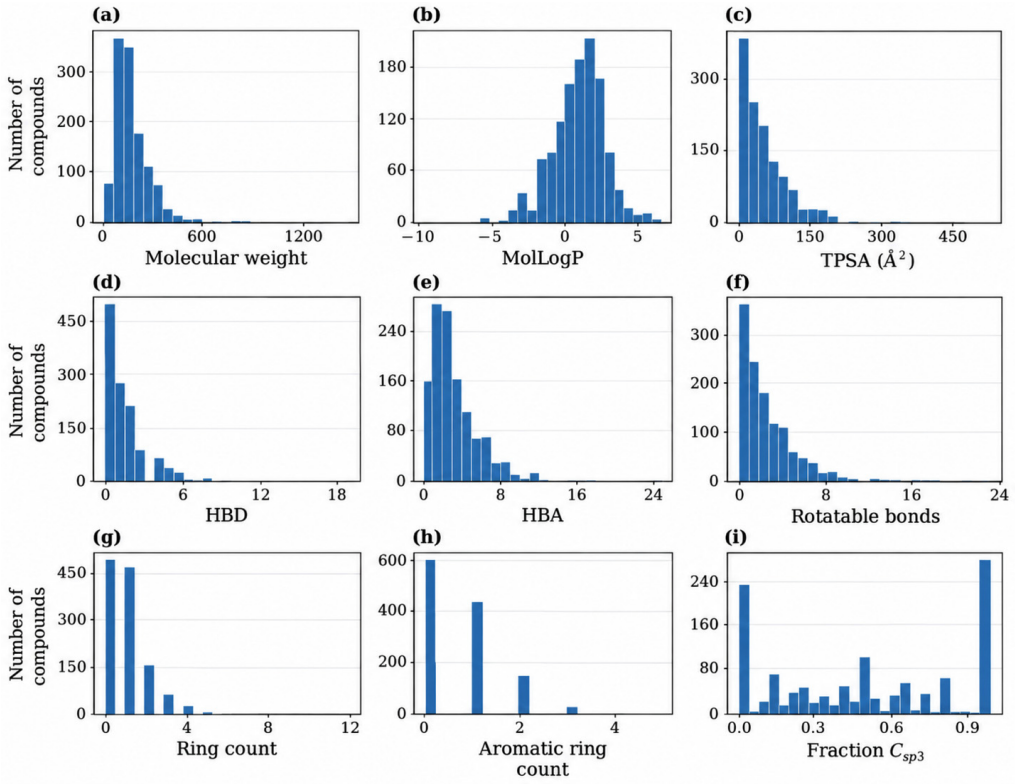



**Figure 2.** Distribution of key molecular physicochemical properties of compounds in the final dataset, including molecular weight (MW) (**a**), octanol–water partition coefficient (MolLogP) (**b**), topological polar surface area (TPSA) (**c**), hydrogen-bond donors (HBD) (**d**), hydrogen-bond acceptors (HBA) (**e**), number of rotatable bonds (**f**), ring count (**g**), aromatic ring count (**h**), and FractionCSP3 (**i**).

### *3.2. Morgan Fingerprint Length Selection*

The influence of Morgan fingerprint length was evaluated using five-fold CV within the training set. Figure 3 presents the mean validation $R^2$ for RF, XGBoost, and SVR at fingerprint lengths of 512, 1024, 2048, 3072, 4096, and 8192 bits. Increasing the length from 512 to 2048 bits improved performance for all three algorithms. Beyond 2048 bits, RF exhibited modest incremental improvements, whereas XGBoost and SVR showed fluctuating performance. At 2048 bits, the mean validation $R^2$ values were 0.533, 0.541, and 0.488 for RF, XGBoost, and SVR, respectively. Although the 8192-bit fingerprint yielded the highest mean validation $R^2$, the corresponding gains in performance relative to the 2048-bit setting were approximately only 0.011, 0.027, and 0.005. The 2048-bit representation was therefore chosen for subsequent model comparisons as a compromise between fingerprint dimensionality and predictive performance. The 2048-bit configuration has also been adopted in previous studies using Morgan fingerprints [48].

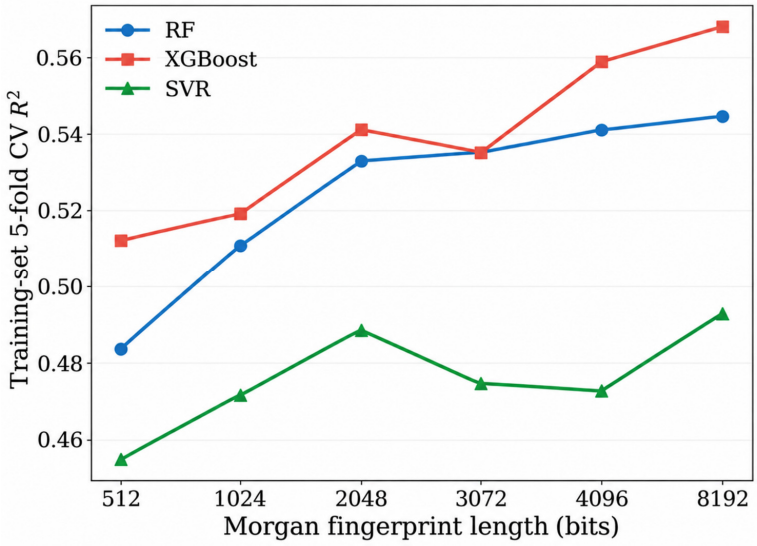



**Figure 3.** Predictive performance of ML algorithms with different Morgan fingerprint lengths.

### *3.3. Model Performances and Global Model Selection*

The 2048-bit Morgan fingerprints, MACCS keys, and RDKit 2D descriptors were combined with RF, XGBoost, and SVR algorithms to generate nine distinct models. In the repeated training-set analysis, all models were evaluated using identical five-fold CV splits. The cross-validation results are summarized in Table 1. Among all models, MACCS–RF yielded the lowest mean validation RMSE (0.0910 ± 0.0051) and the highest mean validation R2 (0.6198 ± 0.0322; values are reported as mean ± standard deviation). RDKit2D–XGBoost showed comparable performance, with a mean validation RMSE of 0.0916 ± 0.0086 and an $R^2$ of 0.6147 ± 0.0587. Although RDKit2D–XGBoost produced a slightly lower mean MAE, MACCS-RF showed smaller fold-to-fold variability for both RMSE and $R^2$. MACCS-RF was therefore selected based on the criterion of minimum mean validation RMSE. The minor performance gap between the two models should not be regarded as evidence of statistically significant superiority of one model over the other.

**Table 1.** Five-fold cross-validation performance of the nine models within the training set.

| Model | $R^2$ (±SD) | RMSE (±SD) | MAE (±SD) | Pearson's r (±SD) |
|---|---|---|---|---|
| MACCS–RF | 0.6198 ± 0.0322 | 0.0910 ± 0.0051 | 0.0634 ± 0.0020 | 0.7906 ± 0.0194 |
| RDKit2D–XGBoost | 0.6147 ± 0.0587 | 0.0916 ± 0.0086 | 0.0630 ± 0.0051 | 0.7863 ± 0.0374 |
| RDKit2D–RF | 0.6091 ± 0.0519 | 0.0923 ± 0.0078 | 0.0640 ± 0.0043 | 0.7827 ± 0.0337 |
| MACCS–XGBoost | 0.5978 ± 0.0441 | 0.0936 ± 0.0073 | 0.0658 ± 0.0032 | 0.7834 ± 0.0239 |
| MACCS–SVR | 0.5719 ± 0.0534 | 0.0966 ± 0.0084 | 0.0684 ± 0.0025 | 0.7650 ± 0.0275 |
| RDKit2D–SVR | 0.5434 ± 0.0462 | 0.0998 ± 0.0074 | 0.0700 ± 0.0056 | 0.7542 ± 0.0321 |
| Morgan–XGBoost | 0.5410 ± 0.0532 | 0.0999 ± 0.0065 | 0.0706 ± 0.0030 | 0.7408 ± 0.0322 |
| Morgan–RF | 0.5329 ± 0.0630 | 0.1008 ± 0.0083 | 0.0694 ± 0.0033 | 0.7328 ± 0.0442 |
| Morgan–SVR | 0.4881 ± 0.0482 | 0.1056 ± 0.0072 | 0.0752 ± 0.0019 | 0.7054 ± 0.0367 |

Note: SD denotes the standard deviation across five validation folds.

The overall training set and held-out test-set results are both summarized in Table 2 for the nine models. MACCS-RF achieved the highest test-set $R^2$ (0.7001), the lowest RMSE (0.0815) and MAE (0.0567), and the highest Pearson's r (0.8377), showing superior predictive performance as the best global model for this dataset. MACCS–XGBoost instead of RDKit2D-XGBoost, ranked second with respect to test-set RMSE, yielding an $R^2$ of 0.6784 and an RMSE of 0.0844. Performances of RDKit2D-XGBoost and RDKit2D–RF were also competitive, with test-set $R^2$ values of 0.6590 and 0.6559, respectively. In contrast, the

Morgan-based models delivered test-set $R^2$ values ranging from 0.3640 to 0.5276 under the present modeling settings. Scatter plots of the observations versus predictions for all nine models on the held-out test-set are provided in Figure S2, which supplement the numerical performance metrics in Table 2.

**Table 2.** Training set and held-out test-set performance of the nine models.

| Representation | Algorithm | Training $R^2$ | Training RMSE | Test $R^2$ | Test RMSE | Test MAE | Test Pearson′ r |
|---|---|---|---|---|---|---|---|
| RDKit 2D | XGBoost | 0.9818 | 0.0201 | 0.6590 | 0.0869 | 0.0595 | 0.8156 |
| RDKit 2D | RF | 0.9470 | 0.0343 | 0.6559 | 0.0873 | 0.0611 | 0.8174 |
| RDKit 2D | SVR | 0.9275 | 0.0402 | 0.4544 | 0.1099 | 0.0785 | 0.6954 |
| MACCS | RF | 0.9365 | 0.0376 | 0.7001 | 0.0815 | 0.0567 | 0.8377 |
| MACCS | XGBoost | 0.9739 | 0.0241 | 0.6784 | 0.0844 | 0.0593 | 0.8238 |
| MACCS | SVR | 0.9665 | 0.0273 | 0.5830 | 0.0961 | 0.0682 | 0.7705 |
| Morgan | XGBoost | 0.9471 | 0.0343 | 0.5276 | 0.1023 | 0.0702 | 0.7287 |
| Morgan | RF | 0.9336 | 0.0384 | 0.5140 | 0.1037 | 0.0712 | 0.7212 |
| Morgan | SVR | 0.9934 | 0.0121 | 0.3640 | 0.1186 | 0.0854 | 0.6100 |

The difference between training-set and test-set performance highlights the importance of evaluating model generalization. Morgan–SVR achieved a training $R^2$ of 0.9934 but a test-set $R^2$ of only 0.3640, whereas MACCS–RF showed better test-set performance despite its slightly lower training $R^2$ of 0.9365. The generalization gaps of all nine models are presented in Figure S3, with MACCS–RF showing the smallest gap.

The performance of the ML models also varied markedly with the choice of molecular representations. RF performed best with MACCS keys, while XGBoost achieved the lowest test-set RMSE among models using RDKit 2D descriptors and Morgan fingerprints. The compact, predefined MACCS structural keys can provide useful predictive information while limiting model complexity [33]. The effectiveness of a molecular representation also depends on the learning algorithm and dataset characteristics.

### *3.4. Error Heterogeneity and Local Optimization Across Structural Classes*

Although MACCS–RF attained the best overall performance, its accuracy may vary across structural classes. The OOF predictions of the global MACCS-RF model actually behaved distinctly across different structural classes. The definitions and compound counts of selected representative structural classes in the full dataset, training set, and held-out test set are provided in Table S2. Application of the predefined OOF-based screening criteria identified sulfur-containing compounds, carboxylic acids/carboxylates, amides, and carbonyl-containing compounds as the major high-error classes before examination of the held-out test. These predefined classes were subsequently evaluated on the held-out test set (Figure 4). The corresponding held-out RMSE values were 0.1064 for sulfur-containing compounds, 0.1063 for carboxylic acids/carboxylates, 0.1046 for amides, and 0.0931 for carbonyl-containing compounds, compared with the overall held-out RMSE of 0.0815. Indeed, the relatively poor performance of the model for these classes on training compounds persisted for unseen test compounds. Besides the four classes, nitriles and esters also exhibited relatively high RMSE values on the held-out test set (Figure 4), though their test-set compound counts were only 6 and 13, respectively.

Since the candidate high-error classes and routing rules were already fixed via training-set OOF analysis, these test-set results were not used to add new classes post hoc to the local modeling strategy. For the four pre-screened high-error classes, MR-ML combinations and hyperparameters were optimized exclusively using the corresponding training compounds and five-fold CV. The corresponding RMSE values for the nine candidate models within each structural class are listed in Table S3 and shown in Figure S1a. The selected class-specific

models and their key hyperparameters are summarized in Table S4. In general, the held-out evaluation revealed that local modeling did not necessarily improve performance for these high-error classes (Table 3). For carboxylic acids/carboxylates, the selected RDKit2D–SVR local model even raised the RMSE from 0.1063 to 0.1274. For carbonyl-containing compounds, RMSE also increased from 0.0931 to 0.0949. The selected RDKit2D–XGBoost model lowered the RMSE for amides from 0.1046 to 0.0989, yet no improvement was found in MAE. In contrast, the RDKit2D-SVR local model established for sulfur-containing compounds reduced the RMSE from 0.1064 to 0.0954 (10.35% drop) and MAE from 0.0725 to 0.0625 (13.88% drop), delivering the most substantial held-out improvement among the four classes.

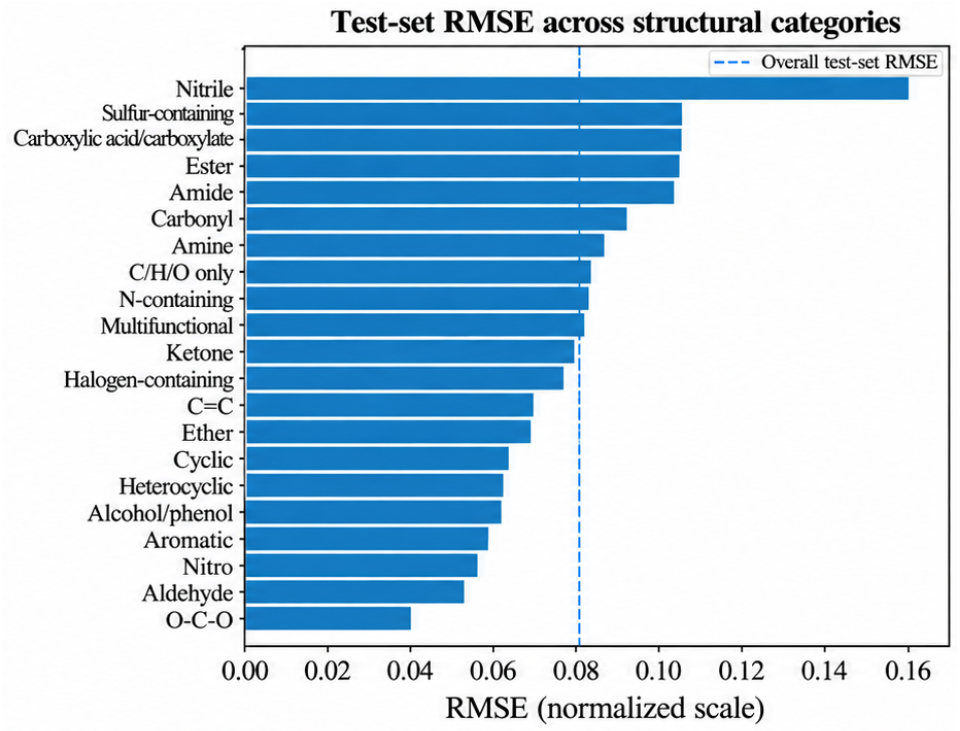



**Figure 4.** Held-out test-set RMSE of the global MACCS-RF model across different structural classes. The dashed line represents the overall RMSE. The highlighted four high-error classes were pre-screened using training set OOF predictions.

**Table 3.** Held-out test-set performance of the global model (MACCS-RF) and class-specific models selected exclusively by five-fold cross-validation within the training set.

| Class | Test Compound Counts | Global Model RMSE | Local Model | Local Model RMSE | Relative Change | Interpretation |
|---|---|---|---|---|---|---|
| Carboxylic acid/ carboxylate | 37 | 0.1063 | RDKit2D–SVR | 0.1274 | −19.89% | No improvement |
| Sulfur-containing compounds | 28 | 0.1064 | RDKit2D–SVR | 0.0954 | +10.35% | improved |
| Amides | 28 | 0.1046 | RDKit2D-XGBoost | 0.0989 | +5.41% | improved |
| Carbonyl-containing compounds | 85 | 0.0931 | RDKit2D–SVR | 0.0949 | −1.94% | No improvement |

To further assess the statistical robustness of the performance for sulfur-containing compounds, paired bootstrap resampling was adopted to quantify the uncertainty in the observed error reduction. This analysis was only used to characterize the uncertainty of the final held-out performance difference, rather than to select the sulfur-containing class or its local model. The RDKit2D–SVR local model indeed yielded a high bootstrap probability

of performance gain. However, the lower bound of the 95% confidence interval was near zero, indicating that this improvement was not statistically robust. The paired-bootstrap percentage changes in RMSE and MAE, alongside the 95% confidence intervals for RMSE change, are reported in Table S4; the RMSE changes and their confidence intervals are illustrated in Figure S1b. The finding is best interpreted as a favorable trend supported by held-out internal validation, while independent external validation remains necessary to assess model transferability beyond the current dataset.

Next, we integrated the local model into the global prediction framework to assess the overall effect. The hierarchical routing rule was applied without modification to the 246 held-out test compounds. In total, 28 sulfur-containing compounds were routed to the RDKit2D–SVR local model, whereas the remaining 218 compounds were predicted using the global MACCS–RF model. The hierarchical strategy increased the held-out test-set $R^2$ from 0.7001 to 0.7115 and reduced RMSE from 0.0815 to 0.0799. The MAE decreased from 0.0567 to 0.0556, and Pearson's *r* increased from 0.8377 to 0.8449 (Figure 5). The overall gain was modest and should therefore be treated as targeted correction for a locally underperforming region of chemical space, rather than a universal improvement across the entire chemical space.

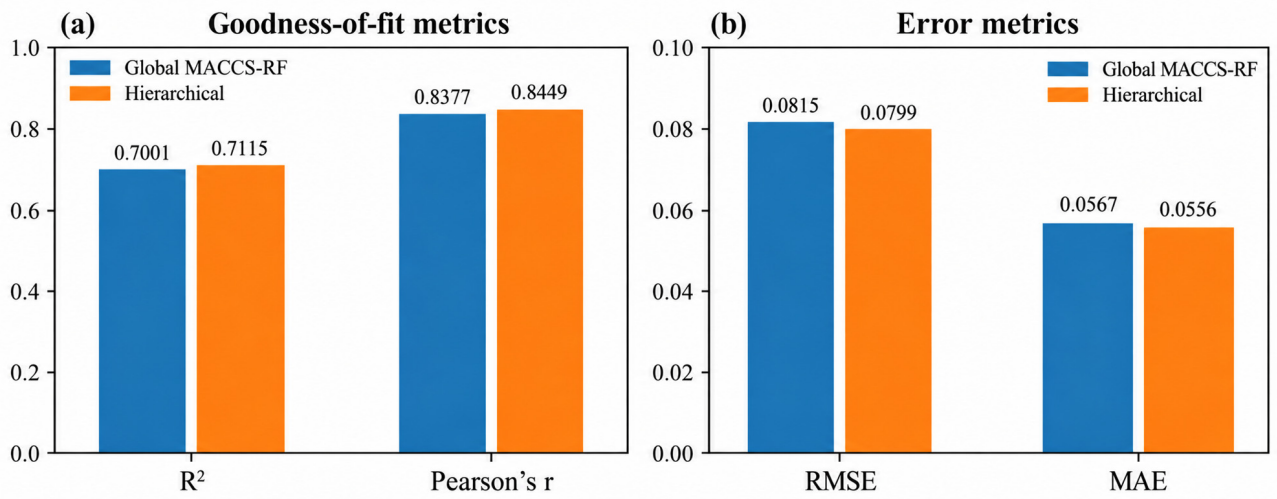



**Figure 5.** Held-out test-set performance comparison between the global MACCS-RF model and the hybrid model refined by the structure-guided hierarchical strategy. Goodness of fit metrics ($R^2$ and Pearson's r) (**a**). Error metrics (RMSE and MAE) (**b**).

Overall, local modeling performance depends on sample size, target distribution, structural heterogeneity and variable structure–reactivity relationships [23]. They support fixing class-selection and routing rules using training-set OOF analysis, instead of choosing local models retrospectively based on held-out test performance. Moreover, sulfur-containing compounds include chemically diverse subtypes with different sulfur oxidation states and functional groups. These differences may affect their reactions with •OH and help explain the relatively large errors. Accordingly, the RDKit2D–SVR local model represents only a data-driven predictive refinement rather than evidence for a universal, unified structure–activity relationship specific to sulfur-containing compounds.

### *3.5. SHAP Analysis*

Since MACCS-RF was selected as the global model via training-set CV, SHAP analysis was conducted on the final MACCS-RF model refitted using the complete training set. The top 10 keys ranked by SHAP importance are shown in Figure 6, and their detailed structural interpretations (representative atom groups, structural types, and predominant contribu-

tion directions) are provided in Table 4. Note that positive SHAP values indicate that the corresponding structural descriptors contribute to higher predicted aqueous •OH reaction rate constants. The SHAP beeswarm plot identified MACCS_162 and MACCS_165 as the two most influential descriptors. High values of these descriptors were mostly associated with positive SHAP values, indicating that aromatic and cyclic structural motifs generally raised the predicted $k_{OH}$ values. MACCS_99 and MACCS_88 also showed predominantly positive contributions and corresponded to C=C-containing and sulfur-containing structures, respectively. In contrast, MACCS_134, MACCS_107, MACCS_123, and MACCS_98 had mainly negative values.

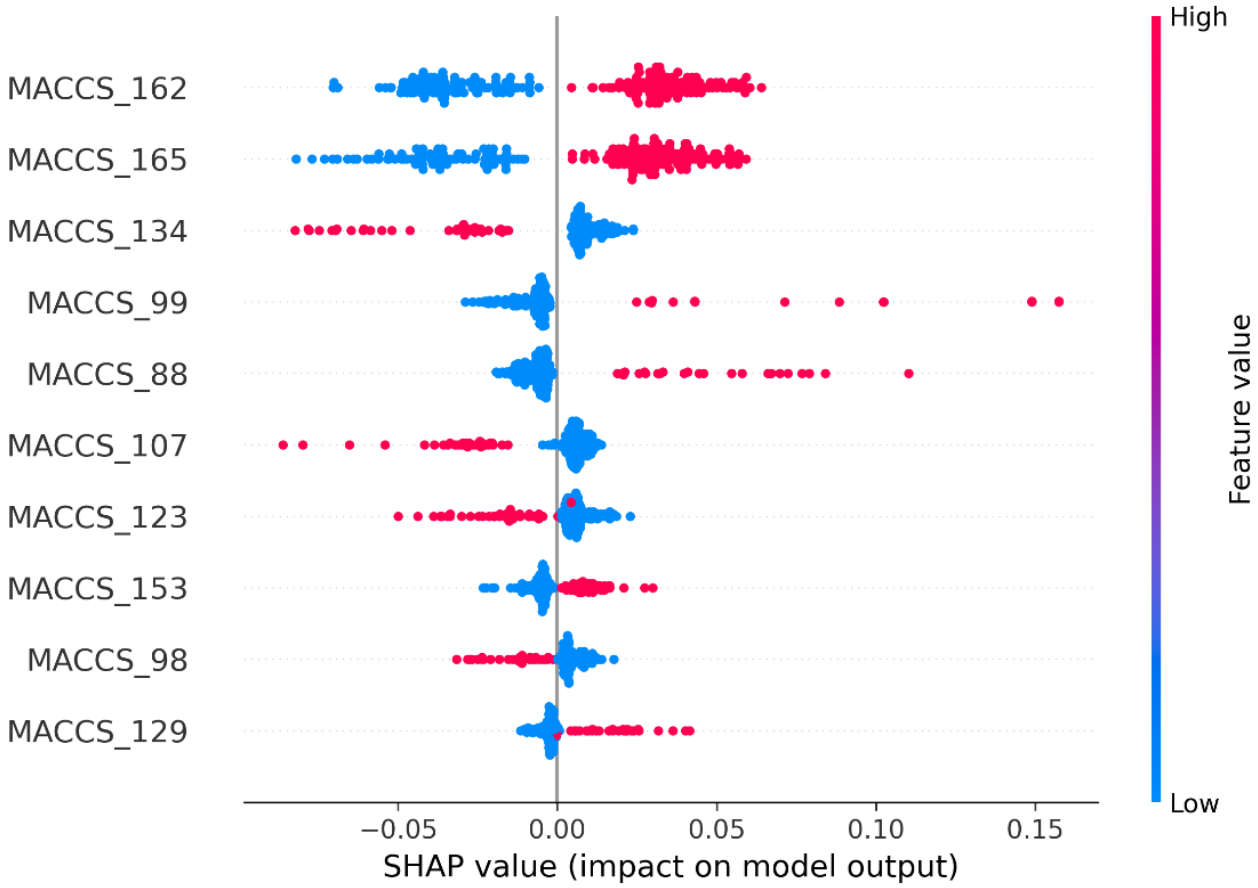



**Figure 6.** SHAP beeswarm analysis of the MACCS–RF model. Colors indicate feature values (red: high; blue: low), and the x-axis represents SHAP values.

The structural associations observed in the SHAP analysis are chemically plausible in the context of established •OH pathways, including electrophilic addition, hydrogen-atom abstraction, and electron transfer [49]. Prior studies on ionizable compounds further demonstrate that protonation status can significantly alter •OH reactivity [24,25]. Nevertheless, each MACCS key indicates merely the presence or absence of a predefined structure motif and cannot fully capture substitution positions, local electronic environment, protonation states, or solvation effects, etc. The SHAP results should be treated as model-derived feature importance, not direct mechanistic evidence of reactions [42,44]. Overall, these findings show that molecular structures such as aromatic rings, C=C bonds, and sulfur-containing groups are important for predicting aqueous •OH reactivity.

**Table 4.** Top 10 important MACCS keys identified by SHAP analysis of the MACCS-RF model, with their representative atom groups, structural features, and predominant contribution directions.

| Feature | Atom Group | Effect on Reactivity | Structure | Feature | Atom Group | Effect on Reactivity | Structure |
|---|---|---|---|---|---|---|---|
| 162 |  | Increase | AROMATIC | 107 |  | Decrease | XA(A)A |
| 165 |  | Increase | RING | 123 |  | Decrease | OCO |
| 134 |  | Decrease | X(HALOGEN) | 153 |  | Increase | QCH2A |
| 99 |  | Increase | C=C | 98 |  | Decrease | QAAAAA@1 |
| 88 |  | Increase | S | 129 |  | Increase | ACH2AACH2A |

Note: The symbol "@" in the SMARTS pattern denotes a ring bond and is part of the original MACCS structural definition.

### *3.6. Application Domain of the Prediction Framework*

The Tanimoto similarity threshold was used to define the applicability domain (AD) boundary by quantifying molecular structural similarity. As the predefined threshold increased, the number of held-out compounds falling within the AD decreased. For the Morgan case, the decline of in-domain compounds was continuous, whereas the RMSE varied irregularly for thresholds from 0.3 to 0.5 (Figure 7a). At a threshold of 0.50, 138 held-out test samples lay within the AD, corresponding to a coverage of 56.10% across all compounds and an RMSE of 0.0969. For the MACCS case, the number of in-domain compounds remained at 245 across thresholds up to 0.4, while the value dropped to 242 when the threshold reached 0.5, corresponding to a coverage of 98.37% and an RMSE of 0.0830 (Figure 7b). Acetylene was the only test compound outside the MACCS-defined AD at thresholds of 0.10–0.40. Likely, its extremely simple structure activates only one MACCS key, corresponding to a carbon–carbon triple bond. Training compounds also with this key generally contain additional structural features, which lower their Tanimoto similarity to acetylene and may explain its low maximum similarity of 0.0769. For MACCS, increasing the threshold from 0.40 to 0.50 removed only three additional compounds from the AD, and a small decrease in RMSE. These results show that most held-out test compounds had MACCS similarity greater than 0.50 relative to at least one training-set compound. However, since Morgan and MACCS descriptors produce different similarity distributions, the same numerical threshold does not define an equivalent boundary in chemical space. For Morgan fingerprints, increasing the threshold excluded many test compounds in the AD, but RMSE did not decrease significantly. Thus, a stricter similarity threshold did not necessarily improve prediction accuracy for the retained compounds.

The Williams plot further evaluated the RDKit 2D-based model using leverage values and standardized residuals (Figure 7c). The warning leverage threshold ($h^*$) was calculated as 0.095 using the equation given in Section 2.6. Most training and held-out test compounds had leverage values below this threshold (Figure 7c). A small subset of potential extrapolation compounds had $h_i > 0.095$, and several prediction outliers had standardized residuals beyond ±3. Some test compounds exhibited both high leverage and large residuals. Accordingly, predictions for unknown compounds are better with structural similarity or leverage metrics; prediction reliability sometimes cannot be judged solely from the overall $R^2$.

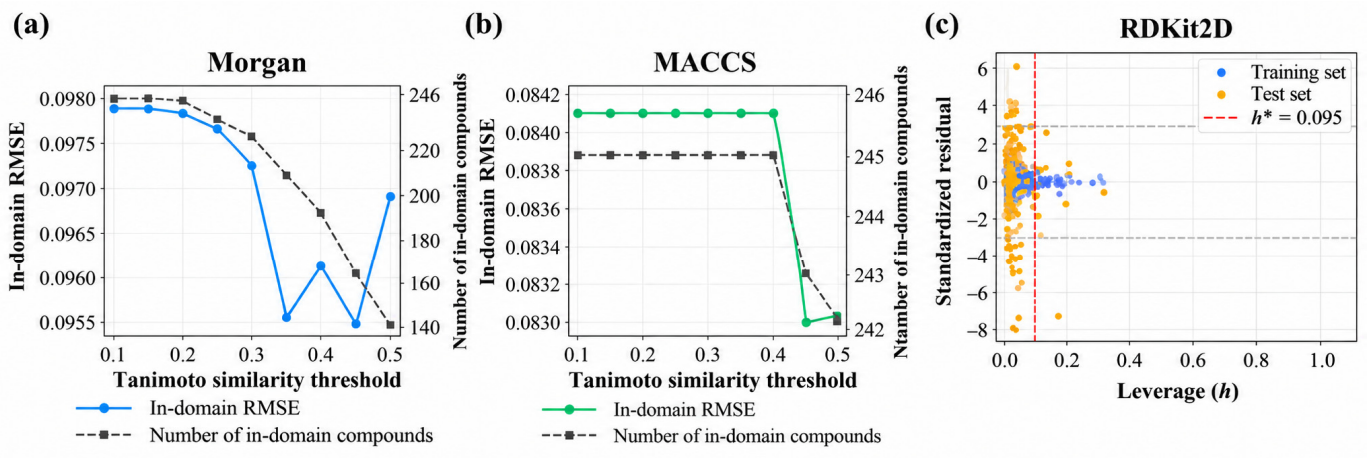



**Figure 7.** Number of in-domain compounds and RMSE across different Tanimoto similarity thresholds for (**a**) Morgan-based and (**b**) MACCS-based models; (**c**) Williams plot for the RDKit2D-based model. The red dashed line denotes $h^* = 0.095$, and the gray dashed lines denote standardized residuals of ±3.

Generally, the Tanimoto-based AD assesses whether a query molecule has structurally similar counterparts in the training set, whereas the Williams plot identifies high-leverage compounds and response outliers in continuous descriptor space. These approaches capture complementary facets of model applicability [46,47]. Overall, SHAP analysis helps identify the structural features influencing model predictions, while AD assessment helps identify compounds for which predictions should be noted.

### *3.7. Comparison with Previous Studies*

Some previous studies have used QSPR and ML methods to predict aqueous-phase •OH reaction rate constants. Gupta and Basant [5] developed a QSPR model based on 995 organic compounds and five molecular descriptors. Their model achieved an $R^2$ of 0.925 on the validation set. Zhong et al. [10] used deep neural networks (DNN), XGBoost, and SHAP analysis to study the •OH reactivity of 1089 organic compounds. Their models achieved test-set $R^2$ values of 0.60–0.71. Sanches-Neto et al. [11] developed the pySiRC framework by combining molecular fingerprints and ML algorithms. Their models also included SHAP interpretation and applicability domain assessment, with external test-set $R^2$ values of 0.639–0.823 for •OH reactions.

In comparison, the present study used a larger dataset with 1229 structurally unique organic compounds and evaluated nine combinations of molecular representations and ML algorithms. Our model in general has predictive performance within the range reported in previous machine-learning studies. However, this does not necessarily indicate better predictive accuracy, as the studies used different datasets, target transformations, and validation methods. In particular, RMSE values reported on different target scales cannot be directly compared. Compared with previous studies, the present work places more emphasis on prediction errors across different classes. We identified structural classes with relatively large errors and evaluated local models for these compounds. We also combined SHAP analysis and AD assessment to better understand model predictions and their reliability. This approach provides additional information on where the model performs well and where further improvement is needed.

## 4. Implications and Limitations

This work integrated multiple MRs and diverse ML algorithms within a unified modeling pipeline governed by consistent training protocols, together with structural-class error

diagnosis and targeted local modeling, while covering a broad chemical space of target species. This sophisticated workflow provides a robust framework for identifying reliable chemical-space regions and pinpointing localized model weaknesses. The predicted rate constants can help assess the reactivity of organic compounds in atmospheric droplets and aquatic environments [1,2,21]. For compounds without experimental kinetic data, the model provides initial estimates of their aqueous-phase reactions with •OH, which may help researchers identify compounds that require further investigation. These predictions can also provide useful kinetic information for future studies of organic compound transformation and environmental persistence [50]. However, the reaction rate constant alone cannot determine environmental lifetime, which also depends on •OH concentrations and other removal processes [50]. Together with the AD assessment, the model can help identify predictions that should be interpreted with caution and compounds that may require further experimental study.

Several limitations, however, should be noted. First, although kinetic data measured at room temperature and under generally consistent experimental conditions were preferentially adopted here, pH value, ionic strength, solution composition, and experimental procedures were practically difficult to fully standardize across all data records. Such experimental heterogeneity may particularly affect the model performance for ionizable compounds [24,25]. Second, although fingerprint-length selection, global model selection, structural class identification, class-specific model optimization, and hierarchical routing strategies were all determined via training-set CV and OOF predictions, the final model evaluation was conducted on a single random data split from the unified curated dataset. Validation using alternative partitioning approaches, including chemical space segmentation, scaffold-based splitting, and similarity/cluster-based division, or validation using complete independent external datasets, could further enhance the assessment of model transferability to structurally novel compounds [23,51,52]. Third, the overall predictive improvement brought by hierarchical modeling was moderate, so the local optimization only achieved targeted error correction for a poorly predicted specific group of compounds, rather than delivering universal performance enhancement across the entire chemical space. Finally, both MACCS structural keys and SHAP analysis only provided statistically derived structural interpretations, without offering direct mechanistic reaction evidence. Further experimental kinetic measurements or quantum chemical calculations are likely helpful to systematically verify the contributions of the key structural descriptors identified in this work.

## 5. Conclusions

We developed a machine-learning framework for predicting aqueous •OH reaction rate constants using 1229 structurally unique organic compounds. Nine combinations of three MRs (Morgan fingerprints, MACCS keys, and RDKit 2D descriptors) coupled with three ML algorithms (RF, XGBoost, and SVR) were compared under a training-only five-fold CV and OOF selection framework, with the held-out test set reserved for final evaluation. The MACCS-RF model was selected as the global model based on the lowest mean validation RMSE obtained from five-fold CV on the training set. Training-set OOF analysis identified sulfur-containing compounds, carboxylic acids/carboxylates, amides, and carbonyl-containing compounds as structural classes with large prediction errors. After class-specific models and routing rules were fixed using only training data, assigning the predefined sulfur-containing subset to the RDKit2D-SVR model increased the overall held-out $R^2$ from 0.7001 to 0.7115 and reduced RMSE from 0.0815 to 0.0799. This improvement was modest and targeted to the pre-screened sulfur-containing species, not a general improvement across the whole chemical space. Subsequent SHAP analysis identified aromatic,

cyclic, C=C-related, and sulfur-containing structural keys as predominantly positive contributors, whereas halogen-related and O–C–O motifs showed mainly negative contributions. Finally, Tanimoto-similarity-based AD assessment revealed that the MACCS model achieved far higher test-set coverage (98.37%) than the Morgan model (56.10%) at the same threshold of 0.50, while Williams-plot diagnosis identified high-leverage and outlier compounds for the RDKit 2D model, demonstrating that global statistical metrics alone are insufficient to judge prediction reliability. Overall, this study shows the value of combining model comparison, structural-class error analysis, and AD assessment for predicting aqueous •OH reaction rate constants. Future work may improve model performance by using more experimental data, additional molecular descriptors, and independent external validation.

**Supplementary Materials:** The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/toxics14111017/s1: Table S1: Data curation of aqueous •OH reaction rate constants; Table S2: Structural classes used for error diagnosis; Table S3: Five-fold cross-validation RMSE of candidate class-specific models; Table S4: Selected class-specific models and paired bootstrap stability assessment; Figure S1:Training-only selection and stability evaluation of class-specific models; Figure S2: Observed versus predicted values for the nine models on the held-out test set; Figure S3: Generalization gap across the nine candidate models.

**Author Contributions:** Conceptualization, X.G. and T.S.; methodology, T.S. and X.G.; software, T.S.; validation, T.S. and Z.Z.; formal analysis, T.S. and Z.Z.; investigation, T.S., Z.Z. and J.Y.; resources, X.G. and J.Y.; data curation, T.S., Z.Z. and N.R.; writing—original draft preparation, T.S.; writing—review and editing, X.G., Y.W., Y.Z. and T.S.; visualization, T.S.; supervision, X.G. and Y.W.; project administration, X.G.; funding acquisition, X.G. All authors have read and agreed to the published version of the manuscript.

**Funding:** National Natural Science Foundation of China (22361162668).

**Data Availability Statement:** The data presented in this article are available on request from the corresponding author.

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

## References

1. Singh, A.; Majumder, A.; Saidulu, D.; Bhattacharya, A.; Bhatnagar, A.; Gupta, A.K. Oxidative treatment of micropollutants present in wastewater: A special emphasis on transformation products, their toxicity, detection, and field-scale investigations. *J. Environ. Manag.* **2024**, *354*, 120339. [CrossRef] [PubMed]
2. Yang, S.; Sun, S.; Xie, Z.; Dong, Y.; Zhou, P.; Zhang, J.; Xiong, Z.; He, C.S.; Mu, Y.; Lai, B. Comprehensive Insight into the Common Organic Radicals in Advanced Oxidation Processes for Water Decontamination. *Environ. Sci. Technol.* **2024**, *58*, 19571–19583. [CrossRef] [PubMed]
3. Keshavarz, M.H.; Shirazi, Z.; Jafari, M.; Jorfi Shanani, S.M. Predicting aqueous-phase hydroxyl radical reaction kinetics with organic compounds in water, atmosphere, and biological systems. *Process Saf. Environ. Prot.* **2025**, *196*, 106876. [CrossRef]
4. Borhani, T.N.G.; Saniedanesh, M.; Bagheri, M.; Lim, J.S. QSPR prediction of the hydroxyl radical rate constant of water contaminants. *Water Res.* **2016**, *98*, 344–353. [CrossRef] [PubMed]
5. Gupta, S.; Basant, N. Modeling the aqueous phase reactivity of hydroxyl radical towards diverse organic micropollutants: An aid to water decontamination processes. *Chemosphere* **2017**, *185*, 1164–1172. [CrossRef] [PubMed]
6. Yu, X.; Liu, J. Prediction of reaction rate constants of hydroxyl radical with chemicals in water. *Water Environ. Res.* **2021**, *93*, 934–939. [CrossRef] [PubMed]
7. Sanches-Neto, F.O.; Dias-Silva, J.R.; de Oliveira, V.M.; Aquilanti, V.; Carvalho-Silva, V.H. Evaluating and elucidating the reactivity of OH radicals with atmospheric organic pollutants: Reaction kinetics and mechanisms by machine learning. *Atmos. Environ.* **2022**, *275*, 119019. [CrossRef]
8. Zhang, G.; Cheng, Y.; Zhang, G.; Chen, Z.; Jiang, J.; Chen, H.; Liu, Z.; Chen, Y.; Dong, X.; Wang, Z.; et al. Innovative Feature Engineering Driven by Chemical Category in Machine Learning for Optimizing the Prediction of Hydroxyl Radical Reaction Rate Constants. *ACS ES T Eng.* **2025**, *5*, 1995–2006. [CrossRef]

9. Zhong, S.; Hu, J.; Yu, X.; Zhang, H. Molecular image-convolutional neural network (CNN) assisted QSAR models for predicting contaminant reactivity toward OH radicals: Transfer learning, data augmentation and model interpretation. *Chem. Eng. J.* **2021**, *408*, 127998. [CrossRef]
10. Zhong, S.; Zhang, K.; Wang, D.; Zhang, H. Shedding light on "Black Box" machine learning models for predicting the reactivity of HO radicals toward organic compounds. *Chem. Eng. J.* **2021**, *405*, 126627. [CrossRef]
11. Sanches-Neto, F.O.; Dias-Silva, J.R.; Keng Queiroz Junior, L.H.; Carvalho-Silva, V.H. "pySiRC": Machine Learning Combined with Molecular Fingerprints to Predict the Reaction Rate Constant of the Radical-Based Oxidation Processes of Aqueous Organic Contaminants. *Environ. Sci. Technol.* **2021**, *55*, 12437–12448. [CrossRef] [PubMed]
12. Gu, L.; Chen, Z. Predicting reaction rate constants of organic compounds with oxidants in the atmospheric aqueous-phase through multi-task learning. *Atmos. Environ.* **2024**, *337*, 120775. [CrossRef]
13. Liu, Z.; Shang, L.; Huang, K.; Yue, Z.; Han, A.Y.; Wang, D.; Zhang, H. Combining Group Contribution Method and Semisupervised Learning to Build Machine Learning Models for Predicting Hydroxyl Radical Rate Constants of Water Contaminants. *Environ. Sci. Technol.* **2025**, *59*, 857–868. [CrossRef] [PubMed]
14. Awfa, D.; Ateia, M.; Mendoza, D.; Yoshimura, C. Application of Quantitative Structure–Property Relationship Predictive Models to Water Treatment: A Critical Review. *ACS ES T Water* **2021**, *1*, 498–517. [CrossRef]
15. Jin, X.; Peldszus, S.; Huck, P.M. Reaction kinetics of selected micropollutants in ozonation and advanced oxidation processes. *Water Res.* **2012**, *46*, 6519–6530. [CrossRef] [PubMed]
16. Jin, X.; Peldszus, S.; Huck, P.M. Predicting the reaction rate constants of micropollutants with hydroxyl radicals in water using QSPR modeling. *Chemosphere* **2015**, *138*, 1–9. [CrossRef] [PubMed]
17. Mao, D.; Liu, J.; Li, X.; Zuo, M.; Yan, W. Holo-Mol: An explainable hybrid deep learning framework for predicting reactivity of hydroxyl radical to water contaminants based on holographic fused molecular representations. *Chem. Eng. J.* **2024**, *482*, 148652. [CrossRef]
18. Liang, Y.; Huangfu, X.; Huang, R.; Han, Z.; Wu, S.; Wang, J.; Long, X.; Ma, J.; He, Q. Machine learning for predicting halogen radical reactivity toward aqueous organic chemicals. *J. Hazard. Mater.* **2024**, *472*, 134501. [CrossRef] [PubMed]
19. Huang, Z.; Yu, J.; Huang, P.; Han, L.; Zhao, X.; Zhang, Q. Cross-Radical Knowledge Sharing via Graph Neural Networks for Unified Prediction of Aqueous Organic Contaminant Second-Order Radical Reaction Rate Constants. *Environ. Sci. Technol.* **2026**, *60*, 28829–28844. [CrossRef] [PubMed]
20. OECD. *Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models*; OECD Publishing: Paris, France, 2007.
21. Dutschmann, T.M.; Schlenker, V.; Baumann, K. Chemoinformatic regression methods and their applicability domain. *Mol. Inform.* **2024**, *43*, e202400018. [CrossRef] [PubMed]
22. Seal, S.; Mahale, M.; Garcia-Ortegon, M.; Joshi, C.K.; Hosseini-Gerami, L.; Beatson, A.; Greenig, M.; Shekhar, M.; Patra, A.; Weis, C.; et al. Machine Learning for Toxicity Prediction Using Chemical Structures: Pillars for Success in the Real World. *Chem. Res. Toxicol.* **2025**, *38*, 759–807. [CrossRef] [PubMed]
23. van Tilborg, D.; Alenicheva, A.; Grisoni, F. Exposing the Limitations of Molecular Machine Learning with Activity Cliffs. *J. Chem. Inf. Model.* **2022**, *62*, 5938–5951. [CrossRef] [PubMed]
24. Gupta, S.; Basant, N. Modeling the pH and temperature dependence of aqueousphase hydroxyl radical reaction rate constants of organic micropollutants using QSPR approach. *Environ. Sci. Pollut. Res.* **2017**, *24*, 24936–24946. [CrossRef] [PubMed]
25. Luo, X.; Wei, X.; Chen, J.; Xie, Q.; Yang, X.; Peijnenburg, W.J. Rate constants of hydroxyl radicals reaction with different dissociation species of fluoroquinolones and sulfonamides: Combined experimental and QSAR studies. *Water Res.* **2019**, *166*, 115083. [CrossRef] [PubMed]
26. He, L.; Schaefer, T.; Otto, T.; Kroflic, A.; Herrmann, H. Kinetic and Theoretical Study of the Atmospheric Aqueous-Phase Reactions of OH Radicals with Methoxyphenolic Compounds. *J. Phys. Chem. A* **2019**, *123*, 7828–7838. [CrossRef] [PubMed]
27. Herrmann, H. Kinetics of aqueous phase reactions relevant for atmospheric chemistry. *Chem. Rev.* **2003**, *103*, 4691–4716. [CrossRef] [PubMed]
28. Hu, Y.; Zhang, Y.; Wen, L.; Schaefer, T.; Herrmann, H. T- and pH-Dependent Hydroxyl-Radical Reaction Kinetics of Lactic Acid, Glyceric Acid, and Methylmalonic Acid in the Aqueous Phase. *J. Phys. Chem. A* **2025**, *129*, 1983–1992. [CrossRef] [PubMed]
29. Ullah, A.; Afzal, A.; Lim, H.-J. Determination of the Hydroxyl Radical Reaction Rate Constant of Amines in the Aqueous Phase. *ACS Earth Space Chem.* **2024**, *8*, 1394–1405. [CrossRef]
30. Xiao, R.; Ye, T.; Wei, Z.; Luo, S.; Yang, Z.; Spinney, R. Quantitative Structure—Activity Relationship (QSAR) for the Oxidation of Trace Organic Contaminants by Sulfate Radical. *Environ. Sci. Technol.* **2015**, *49*, 13394–13402. [CrossRef] [PubMed]
31. Morgan, H.L. The Generation of a Unique Machine Description for Chemical Structures-A Technique Developed at Chemical Abstracts Service. *J. Chem. Doc.* **1965**, *5*, 107–113. [CrossRef]
32. Rogers, D.; Hahn, M. Extended-connectivity fingerprints. *J. Chem. Inf. Model.* **2010**, *50*, 742–754. [CrossRef] [PubMed]

33. Durant, J.L.; Leland, B.A.; Henry, D.R.; Nourse, J.G. Reoptimization of MDL keys for use in drug discovery. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 1273–1280. [CrossRef] [PubMed]
34. Bento, A.P.; Hersey, A.; Felix, E.; Landrum, G.; Gaulton, A.; Atkinson, F.; Bellis, L.J.; De Veij, M.; Leach, A.R. An open source chemical structure curation pipeline using RDKit. *J. Cheminform.* **2020**, *12*, 51. [CrossRef] [PubMed]
35. RDKit. *RDKit: Open-Source Cheminformatics Software*, Version 2023.09.6; Zenodo: Geneva, Switzerland, 2024. [CrossRef]
36. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
37. Chen, T.; Guestrin, C. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*; Association for Computing Machinery: New York, NY, USA, 2016; pp. 785–794.
38. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [CrossRef]
39. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
40. Correia, J.; Capela, J.; Rocha, M. Deepmol: An automated machine and deep learning framework for computational chemistry. *J. Cheminform.* **2024**, *16*, 136. [CrossRef] [PubMed]
41. Hinkley, D.V. Bootstrap methods. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **1988**, *50*, 321–337. [CrossRef]
42. Lundberg, S.M.; Lee, S.-I. A unified approach to interpreting model predictions. *arXiv* **2017**, arXiv:1705.07874.
43. Lundberg, S.M.; Erion, G.; Chen, H.; DeGrave, A.; Prutkin, J.M.; Nair, B.; Katz, R.; Himmelfarb, J.; Bansal, N.; Lee, S.I. From Local Explanations to Global Understanding with Explainable AI for Trees. *Nat. Mach. Intell.* **2020**, *2*, 56–67. [CrossRef] [PubMed]
44. Wellawatte, G.P.; Gandhi, H.A.; Seshadri, A.; White, A.D. A Perspective on Explanations of Molecular Prediction Models. *J. Chem. Theory Comput.* **2023**, *19*, 2149–2160. [CrossRef] [PubMed]
45. Gramatica, P. Principles of QSAR models validation: Internal and external. *QSAR Comb. Sci.* **2007**, *26*, 694–701. [CrossRef]
46. Sahigara, F.; Mansouri, K.; Ballabio, D.; Mauri, A.; Consonni, V.; Todeschini, R. Comparison of different approaches to define the applicability domain of QSAR models. *Molecules* **2012**, *17*, 4791–4810. [CrossRef] [PubMed]
47. Bajusz, D.; Racz, A.; Heberger, K. Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? *J. Cheminform.* **2015**, *7*, 20. [CrossRef] [PubMed]
48. Zhong, S.; Hu, J.; Fan, X.; Yu, X.; Zhang, H. A deep neural network combined with molecular fingerprints (DNN-MF) to develop predictive models for hydroxyl radical rate constants of water contaminants. *J. Hazard. Mater.* **2020**, *383*, 121141. [CrossRef] [PubMed]
49. Buxton, G.V.; Greenstock, C.L.; Helman, W.P.; Ross, A.B. Critical Review of rate constants for reactions of hydrated electrons, hydrogen atoms and hydroxyl radicals ($\cdot OH/\cdot O^-$ in Aqueous Solution. *J. Phys. Chem. Ref. Data* **1988**, *17*, 513–886. [CrossRef]
50. Bianco, A.; Passananti, M.; Brigante, M.; Mailhot, G. Photochemistry of the Cloud Aqueous Phase: A Review. *Molecules* **2020**, *25*, 423. [CrossRef] [PubMed]
51. Simm, J.; Humbeck, L.; Zalewski, A.; Sturm, N.; Heyndrickx, W.; Moreau, Y.; Beck, B.; Schuffenhauer, A. Splitting chemical structure data sets for federated privacy-preserving machine learning. *J. Cheminform.* **2021**, *13*, 96. [CrossRef] [PubMed]
52. Kretschmer, F.; Seipp, J.; Ludwig, M.; Klau, G.W.; Bocker, S. Coverage bias in small molecule machine learning. *Nat. Commun.* **2025**, *16*, 554. [CrossRef] [PubMed]


**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.