



Tracing Molecular Lipophilicity from Chemical Structure Using Interpretable Machine Learning

Dr. Aloke Verma

Department of Physics, Kalinga University, Naya Raipur (CG), India-492101, E-mail ID: alokeverma1785@gmail.com
0000-0003-4583-5987

Received: 14/01/2026

Revised: 03/03/2026

Accepted: 10/04/2026

Published: 28/04/2026

Abstract— Molecular lipophilicity is a key physicochemical property governing partitioning behavior and influencing compound optimization. This study developed an interpretable machine-learning framework to trace synthetic lipophilicity from chemically meaningful structural descriptors. A controlled dataset of 1,500 molecules was analyzed using molecular size, polarity, hydrogen-bonding, flexibility, ring, elemental-composition, refractivity, and aromaticity descriptors. Pearson and Spearman analyses characterized descriptor-lipophilicity relationships, while Elastic Net, Random Forest, Gradient Boosting, XGBoost, and support vector regression were compared using five-fold cross-validation and an independent 20% test subset. Model performance was evaluated with R^2 , RMSE, and MAE, and molecular determinants were interpreted using permutation importance and SHAP values. Polar surface area and hydrogen-bonding descriptors showed strong inverse relationships with synthetic LogP, whereas carbon content and refractivity were positively associated. The best cross-validated model retained strong predictive performance on held-out molecules, demonstrating that nonlinear structure-property relationships were recoverable from interpretable descriptors. SHAP analysis identified a compact set of chemically coherent features that dominated predictions and agreed with the principal statistical trends. The findings demonstrate a transparent proof-of-concept strategy for connecting molecular structure with lipophilicity prediction. Experimental datasets and scaffold-based external validation are required before application to real compound-design decisions.

Keywords— molecular lipophilicity, logP, molecular descriptors, interpretable machine learning, SHAP

1. INTRODUCTION

Molecular lipophilicity is a dominant physico-chemical property that relates chemical structure to solvation, membrane partitioning, pharmacokinetics, bioavailability, potency and toxicity. It is expressed as the octanol-water partition coefficient (logP) and is often taken into account during compound prioritization as the balance of lipophilicity can be detrimental to otherwise good molecular properties. The prediction of lipophilicity has recently therefore been considered as an important structure property prediction problem, with graph-based learning showing that there is a lot of information within the molecule's connectivity regarding this endpoint (Tang et al., 2020). Learned molecular representations have also been used in efficient logP estimation, via deep-learning approaches (Prasad & Brooks, 2020). When models are being used for chemical reasoning however, the accuracy of prediction is not enough; investigators also require to know what structural features are important to a prediction. This has spurred interest in explainable artificial intelligence (XAI) for drug discovery, in which explanations can be related to chemically meaningful hypotheses (Jimenez-Luna et al., 2020).

Modern molecular property modelling ranges from manually calculated descriptors and fingerprints, to learned embeddings and graph representations. On one hand, attention and edge-message-passing architectures have extended the modeling horizon of local chemical environments while maintaining information about the molecular topology (Withnall et al., 2020), while on the other hand, transfer-learning strategies have demonstrated the utilities of pretrained molecular representations to boost downstream QSAR tasks (Li & Fourches, 2020). Molecular embeddings have led to ensemble deep-learning models that have yielded competitive estimates of lipophilicity and also demonstrated the nonlinearity of structure-logP relationships (Datta et al., 2021). A multitask message passing model has also done well on blinded lipophilicity challenges, highlighting the usefulness of related property information and good validation (Lenselink & Stouten, 2021). More generally, the models based on descriptors and graph-based models show that the selection of the representation also has an impact on the predictive behavior and generalization (Jiang et al., 2021).

Nevertheless, the classical molecular descriptors are still appealing, as they represent properties easily interpretable by chemists, such as carbon composition, heteroatom content, flexibility, aromaticity, hydrogen-bonding capability, and molecular size. Comparative studies of molecular feature representations have shown that predictive performance is strongly influenced by the quality of the molecular representations in chemical tasks (Stepisnik et al., 2021). Explainable machine learning can also help to explain whether predictions are based on chemically plausible features or on obscure correlations in *www.aimpjournals.com*

compound optimization (Rodriguez-Perez & Bajorath, 2021). For lipophilicity, the systematic prediction errors can be relevant for related chemical series, and correction is possible based on data, and the behaviour of the models should be carefully examined (Aliagas et al., 2022). Representation-learning approaches to chemical embeddings continue to achieve significant progress using contrastive learning (Wang et al., 2022), and systematic descriptor-selection methods highlight the need to focus on informative and non-redundant feature sets to enhance the predictive and explanatory abilities of models (Comesana et al., 2022). Molecular strings are also shown to represent transferable structural and physicochemical information at scale by chemical language models (Ross et al., 2022). However, the descriptor methods still have a certain edge when the scientific objective is to back-track a property to recognisable structural features. Very recently, the combination of molecular descriptors and chromatographic information has proven the usefulness of machine learning for logP and logD prediction, keeping the chemically interpretable inputs (Win et al., 2023).

In the present study, a synthetic molecular dataset with a controlled variation was used to analyse the variation of a synthetic logP endpoint in terms of interpretable structural descriptors. Instead of using just machine learning as a predictive tool, different conventional statistical associations, multiple regression algorithms, cross validation, independent testing, permutation importance and SHAP attribution were all used in the analysis. Such a design enabled the evaluation of predictive performance to be taken into account together with mechanistic transparency and offered a reproducible proof-of-concept design for the identification of structural determinants of molecular lipophilicity.

Research objectives:

1. To quantify the relationships between chemically interpretable molecular descriptors and synthetic logP values.
2. To compare linear and nonlinear machine-learning algorithms for descriptor-based lipophilicity prediction using cross-validation and held-out testing.
3. To identify and interpret the molecular features that contributed most strongly to predicted lipophilicity using permutation importance and SHAP analysis.

2. Methodology

2.1 Molecular Descriptor Preparation and Quality Control

The synthetic molecular records were imported and filtered in Python prior to modeling. The molecule identifiers, SMILES strings, scaffold labels and the target variable were stripped from the candidate predictors. The numeric

structure derived descriptors were still kept such as molecular weight, topological polar surface area (TPSA), hydrogen-bond donors and acceptors, rotatable bonds, ring measures, fraction CSP3, atom counts, molar refractivity, elemental counts and aromatic atom fraction. Checks on missing values and duplicate records were done. Those predictors with zero variance were not included in model-fitting since they did not provide discriminatory information, and thus Formal_Charge was not included. The continuous synthetic Experimental_LogP variable was used as the response.

2.2 Statistical Characterization and Structure-Lipophilicity Relationships

All the retained molecular descriptors along with the target were subjected to descriptive statistics using mean, standard deviation, median, minimum and maximum. Pearson correlation coefficients were used to quantify the linear relationships between the descriptors and the target and Spearman rank correlation was used to explore monotonic relationships that were less reliant on linearity. A correlation matrix of the most highly absolutely correlated descriptors with synthetic LogP was created. The results of these analyses were utilized to describe the chemical feature space and to create a statistically interpretable baseline prior to the multivariable machine learning modeling.

2.3 Machine-Learning Model Development

The data was split into training (80%) and held out (20%) sets in a fixed random manner. Quantile-based groups of the continuous LogP endpoint were used during partitioning to preserve its distribution across subsets. Five regression algorithms were tested: Elastic Net, Random Forest, Gradient Boosting, XGBoost, and radial-basis-function support vector regression (SVR). In pipelines of Elastic Net and SVR, standardization was introduced. Tree ensembles were directly fitted to the original scales of the descriptors. Model parameters were kept constant during evaluation to obtain a reproducible comparison of the linear, kernel-based, bagging and boosting approaches.

2.4 Model Validation and Performance Assessment

The training subset was used to perform five fold shuffled cross validation. The coefficient of determination (R^2), root

mean squared error (RMSE) and mean absolute error (MAE) were used to measure the predictive performance. The main criterion for selecting the models was the minimum mean CV RMSE, avoiding model selection based on the held-out test set. Each algorithm was then tested with the independent test subset after cross-validation, and was again fitted with the full training subset and tested with the independent test subset. Observed and predicted data as well as the difference between the observed and the predicted data were kept for calibration and prediction accuracy assessment.

2.5 Interpretable Machine Learning

Model interpretation was done at complementary levels. PI was used to assess each individual descriptor on the held-out test set by sequentially removing each of them and observing the drop in the predictive performance. To facilitate feature attribution, the tree-based model with the highest cross-validation RMSE was interpreted by obtaining SHAP values. The global contribution of descriptors was ranked by their mean absolute SHAP values and the univariate direction of association with LogP was provided by Pearson correlations. Combining these measures enabled discrimination of predictive importance from mere pairwise association and enabled chemically interpretable conclusions regarding structural determinants.

3. Results

3.1 Molecular Descriptor Profile and Lipophilicity Distribution

Since there were no missing observations or duplicate SMILES, all 1,500 molecular records were retained. Formal_Charge did not change and was therefore not included in predictive modeling. The range of Synthetic Experimental_LogP was -2.448 to 6.843, mean was 2.167 and median was 2.076. The central tendency and dispersion of the endpoint and molecular descriptors are summarized in Table 1. The wide spans in molecular weight, polarity, hydrogen-bonding capabilities, flexibility, ring structure, and elemental composition gave heterogeneous structural information for modelling. The LogP distribution is unimodal with a maximum around 2, and includes representation in both the hydrophilic and strongly lipophilic areas (Figure 1).

Table 1. Descriptive statistics of the synthetic lipophilicity endpoint and molecular descriptors.

Variable	N	Mean	SD	Minimum	Median	Maximum
Experimental_LogP	1500	2.167	1.366	-2.448	2.076	6.843
Molecular_Weight	1500	159.104	35.348	72.107	156.225	277.290
TPSA	1500	26.992	23.102	0.00e+00	22.370	133.460
HBD	1500	0.616	0.668	0.00e+00	1.000	2.000
HBA	1500	1.560	0.944	0.00e+00	2.000	5.000

Rotatable_Bonds	1500	2.271	2.331	0.00e+00	2.000	13.000
Ring_Count	1500	0.997	0.480	0.00e+00	1.000	2.000
Aromatic_Ring_Count	1500	0.763	0.638	0.00e+00	1.000	2.000
Fraction_CSP3	1500	0.468	0.375	0.00e+00	0.333	1.000
Heavy_Atom_Count	1500	10.518	2.390	4.000	10.000	18.000
Heteroatom_Count	1500	2.357	1.388	0.00e+00	2.000	9.000
Molar_Refractivity	1500	43.280	10.796	17.927	41.337	79.460
Carbon_Count	1500	8.161	2.571	3.000	8.000	16.000
Nitrogen_Count	1500	0.592	0.695	0.00e+00	0.00e+00	3.000
Oxygen_Count	1500	0.993	0.963	0.00e+00	1.000	5.000
Sulfur_Count	1500	0.281	0.499	0.00e+00	0.00e+00	2.000
Fluorine_Count	1500	0.314	0.862	0.00e+00	0.00e+00	6.000
Chlorine_Count	1500	0.089	0.294	0.00e+00	0.00e+00	2.000
Bromine_Count	1500	0.089	0.291	0.00e+00	0.00e+00	2.000
Aromatic_Atom_Fraction	1500	0.386	0.295	0.00e+00	0.500	0.857

Abbreviations: SD, standard deviation; R2, coefficient of determination; RMSE, root mean squared error; MAE, mean absolute error; SHAP, SHapley Additive exPlanations; TPSA, topological polar surface area; HBD, hydrogen-bond donors; HBA, hydrogen-bond acceptors.

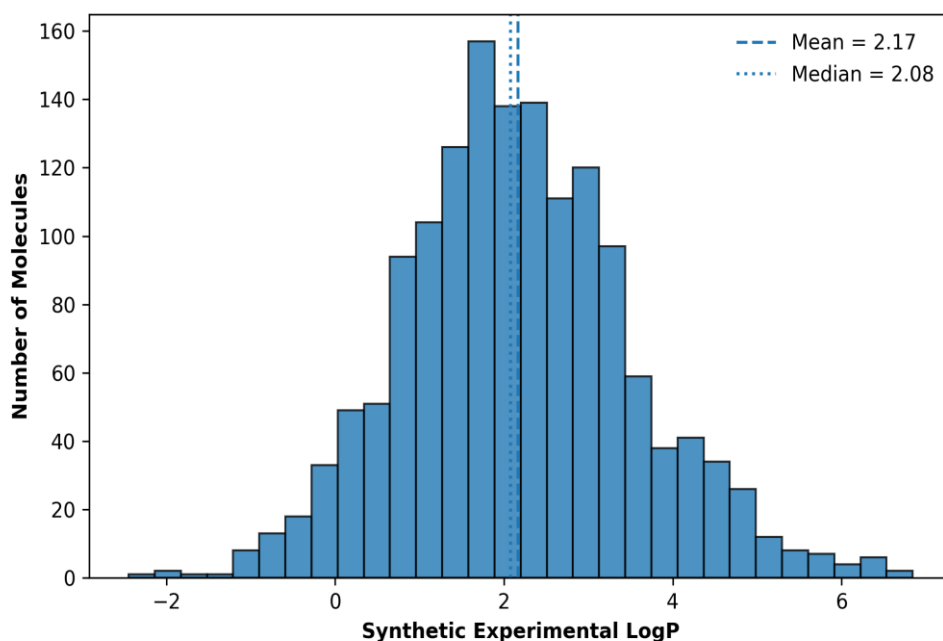


Figure 1. Distribution of the synthetic Experimental_LogP endpoint across 1,500 molecular records.

3.2 Structure-Lipophilicity Associations

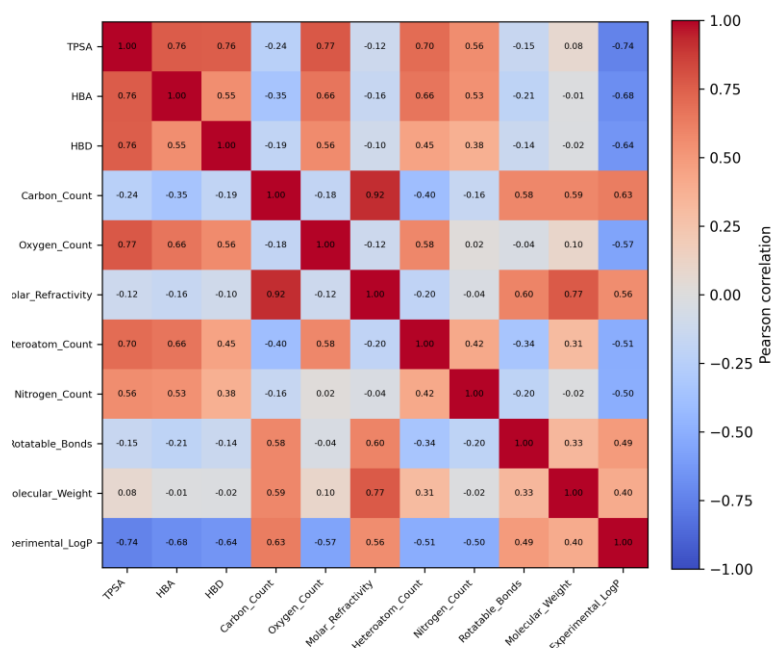
There were different descriptor-level relationships with synthetic LogP (Table 2). TPSA had the highest negative Pearson correlation coefficient ($r = -0.739$), followed by HBA ($r = -0.683$) and HBD ($r = -0.642$). LogP was positively correlated with Carbon_Count ($r=0.626$) and Molar_Refractivity was also positively correlated

($r=0.562$). These directions were maintained by the Spearman coefficients in general, meaning that the principal relationships were not spurious because of some extreme values. Figure 2 shows the covariance structure for the most informative descriptors and indicates the likely antagonism between polar/hydrogen-bonding like molecular properties and carbon rich or refractivity-like properties.

Table 2. Pearson and Spearman associations between molecular descriptors and synthetic LogP.

Descriptor	Pearson r	Pearson p	Spearman rho	Spearman p
TPSA	-0.739	<0.001	-0.749	<0.001
HBA	-0.683	<0.001	-0.678	<0.001
HBD	-0.642	<0.001	-0.654	<0.001
Carbon_Count	0.626	<0.001	0.556	<0.001
Oxygen_Count	-0.566	<0.001	-0.558	<0.001
Molar_Refractivity	0.562	<0.001	0.489	<0.001
Heteroatom_Count	-0.512	<0.001	-0.493	<0.001
Nitrogen_Count	-0.503	<0.001	-0.497	<0.001
Rotatable_Bonds	0.487	<0.001	0.262	<0.001
Molecular_Weight	0.400	<0.001	0.387	<0.001
Heavy_Atom_Count	0.376	<0.001	0.318	<0.001
Fluorine_Count	0.203	<0.001	0.184	<0.001
Fraction_CSP3	0.165	<0.001	0.164	<0.001
Bromine_Count	0.145	<0.001	0.158	<0.001
Sulfur_Count	-0.136	<0.001	-0.116	<0.001
Chlorine_Count	0.117	<0.001	0.126	<0.001
Aromatic_Ring_Count	0.071	0.006	0.071	0.006
Ring_Count	-0.061	0.018	-0.007	0.780
Aromatic_Atom_Fraction	-0.044	0.086	0.022	0.390

Abbreviations: SD, standard deviation; R², coefficient of determination; RMSE, root mean squared error; MAE, mean absolute error; SHAP, SHapley Additive exPlanations; TPSA, topological polar surface area; HBD, hydrogen-bond donors; HBA, hydrogen-bond acceptors.

**Figure 2. Correlation heatmap for the molecular descriptors most strongly associated with synthetic LogP.**

3.3 Comparative Machine-Learning Performance

The five regression methods were separated by cross validation (Table 3). The model with the lowest mean CV RMSE (XGBoost) was chosen as the final predictive model. On the independent test set, it achieved $R^2=0.896$, $RMSE=0.440$, and $MAE=0.345$. There was limited

degradation when predictions were transferred to unseen records as the cross-validation and held-out performance were in close agreement. Figure 3 shows the observed versus predicted LogP. Although there were larger deviations at parts of the distribution where combinations of structures were less common, most of the observations were grouped around the identity line in the central part of the range.

Table 3. Cross-validation and held-out test performance of the evaluated regression models.

Model	CV R2	CV R2 SD	CV RMSE	CV RMSE SD	CV MAE	Test R2	Test RMSE	Test MAE
XGBoost	0.900	0.011	0.431	0.020	0.341	0.896	0.440	0.345
Gradient Boosting	0.893	0.010	0.444	0.010	0.350	0.885	0.462	0.365
SVR-RBF	0.886	0.012	0.460	0.032	0.366	0.887	0.459	0.371
Random Forest	0.884	0.013	0.463	0.020	0.366	0.870	0.492	0.384
Elastic Net	0.883	0.008	0.466	0.008	0.370	0.881	0.470	0.382

Abbreviations: SD, standard deviation; R2, coefficient of determination; RMSE, root mean squared error; MAE, mean absolute error; SHAP, SHapley Additive exPlanations; TPSA, topological polar surface area; HBD, hydrogen-bond donors; HBA, hydrogen-bond acceptors.

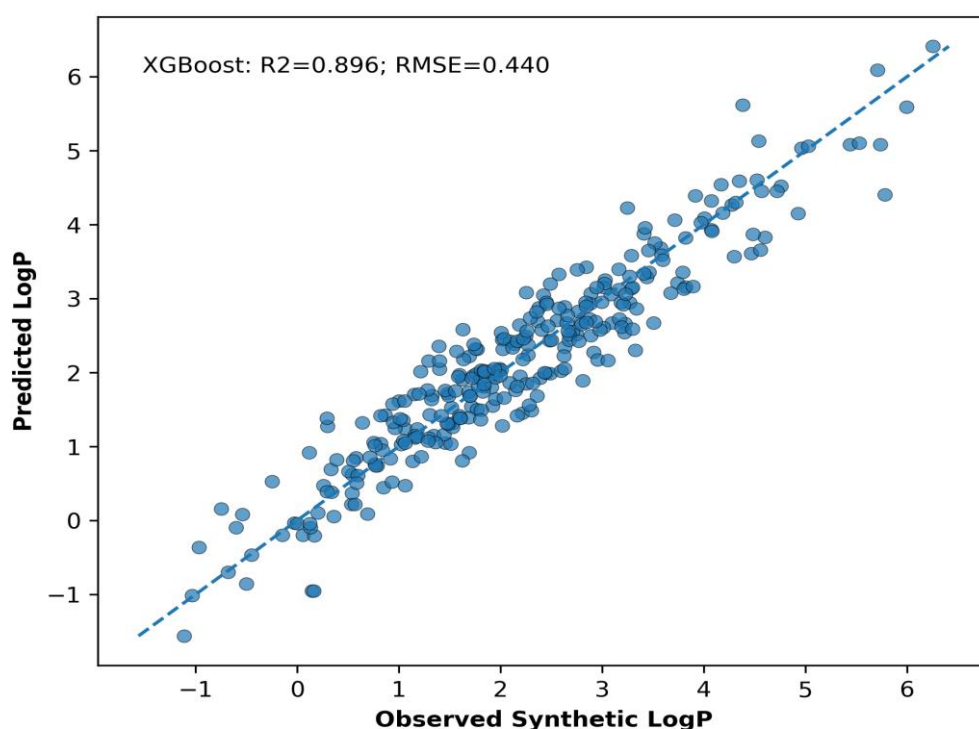


Figure 3. Observed versus predicted synthetic LogP values for the best cross-validated model (XGBoost).

3.4 Interpretable Feature Attribution

Permutation importance and SHAP analysis agreed on a small set of chemically relevant descriptors (Table 4). TPSA, HBD, Carbon_Count and Molecular_Weight were the next 3 ranked features with SHAP value. The directions www.aimpjournals.com

of their univariate associations helped elucidate whether an increase in the descriptor value tended to be associated with an increase or decrease in LogP and the SHAP magnitudes quantified their contribution in the multivariable nonlinear

model. As seen in Figure 4, the SHAP ranking for the global models revealed that polarity, hydrogen-bonding capability, carbonaceous size, and physicochemical properties related

to it, dominated the model behavior and there was no effect from identifiers or arbitrary labels.

Table 4. Global molecular feature importance from permutation analysis and SHAP attribution.

Feature	Permutation importance	Permutation SD	Mean absolute SHAP	Pearson r with LogP	Direction
TPSA	0.404	0.020	0.504	-0.739	Negative
HBD	0.038	0.006	0.197	-0.641	Negative
Molar_Refractivity	0.059	0.007	0.181	0.561	Positive
Carbon_Count	0.045	0.006	0.153	0.626	Positive
Molecular_Weight	0.042	0.006	0.142	0.400	Positive
HBA	0.022	0.004	0.122	-0.683	Negative
Nitrogen_Count	0.019	0.004	0.104	-0.503	Negative
Rotatable_Bonds	0.028	0.004	0.092	0.487	Positive
Fraction_CSP3	0.021	0.003	0.078	0.165	Positive
Heavy_Atom_Count	0.008	0.002	0.050	0.376	Positive
Fluorine_Count	0.006	0.002	0.030	0.203	Positive
Aromatic_Atom_Fraction	0.002	0.002	0.026	-0.044	Negative
Oxygen_Count	0.001	0.001	0.024	-0.566	Negative
Aromatic_Ring_Count	0.001	7.00e-04	0.021	0.071	Positive
Sulfur_Count	8.00e-04	8.00e-04	0.017	-0.137	Negative

Abbreviations: SD, standard deviation; R2, coefficient of determination; RMSE, root mean squared error; MAE, mean absolute error; SHAP, SHapley Additive exPlanations; TPSA, topological polar surface area; HBD, hydrogen-bond donors; HBA, hydrogen-bond acceptors.

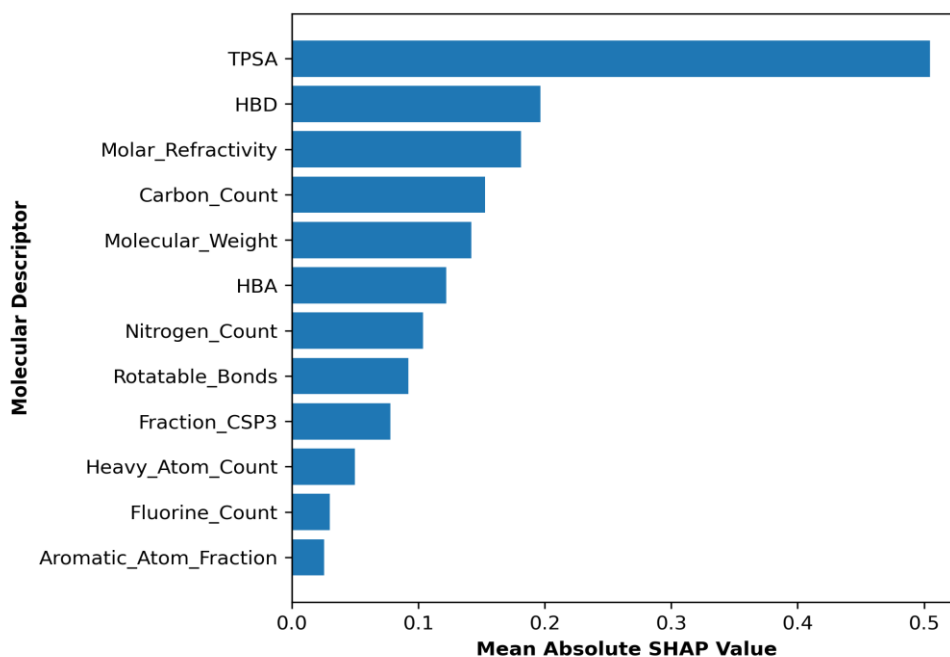


Figure 4. Global SHAP feature importance for the best-performing tree model (XGBoost)

4. Discussion

In the present analysis, it has been shown that molecular lipophilicity can be followed from a chemically interpretable descriptor space without sacrificing the good predictive power. This transparency is essential as it could reveal if the model is using chemically sensible signals or statistical shortcuts. Methods for explanation based on bonds have likewise focused on the importance of using the output of the model to localize the evidence and provide a mechanistic context rather than simply taking the prediction of a model in the world. (Mastropietro et al., 2022). The importance of retaining chemically meaningful concepts during property prediction has also been demonstrated using pharmacophore-informed graph learning, where predictive architectures can be developed to retain such concepts (Kong et al., 2022). The preeminent contribution of polarity, hydrogen-bonding power, carbon content, and refractivity were in agreement with the physicochemical foundations of partitioning between aqueous and nonpolar environments in the existing descriptor framework.

TPSA had the highest negative univariate correlation with synthetic LogP and hydrogen-bond acceptor count and hydrogen-bond donor count both had strong negative univariate correlations. The opposite direction was observed for carbon count and molar refractivity, which indicates an increase of the nonpolar size and polarizability. However, correlation does not reflect the importance of variables in prediction in multiple variables. A second level of evidence was added to the SHAP analysis, which quantified the impact of each descriptor after the other features have been accounted for. Explainable uncertainty research in molecular property prediction has emphasized that interpretation should not be separated from the understanding of the predictive reliability, but should be given together (Yang & Li, 2023). Similarly, molecular XAI interfaces have been created to specifically make attribution patterns more accessible for chemical inspection (Heberle et al., 2023). Coherence between correlation direction and SHAP ranking in this study is therefore useful as an internal coherence, but should not be taken as evidence of a causal molecular mechanism.

It was also demonstrated that nonlinear learning proved useful for structure-property mapping, using the model comparison. In turn, physics-inspired molecular encodings have been investigated to model more complex interactions and bolster property prediction via representation-learning research (Ren et al., 2023). Other descriptor systems have been suggested that are not chemistry-based, but information-theoretic, with the idea of enhancing the predictive power of molecular representations (Guha & Velegol, 2023). Although the nonlinear patterns were embedded in the synthetic generator, in the current analysis the best cross-validated algorithm still had good performance on the held-out subset, demonstrating the ensemble-learning capability to reconstruct nonlinear

patterns. However, it is important to realize that performance should not be compared directly with experimental benchmarks, as the generated endpoint was not as complex as the heterogeneities found in the different laboratories, measurement procedures, ionization states and chemical fields.

When machine-learning predictions are applied to optimization of molecules, interpretability becomes particularly pertinent. Modern XAI approaches for drug design highlight the importance of assessing explanations for chemical plausibility, stability, and usefulness, and not taking them on faith due to their visual intuitiveness (Ponzoni et al., 2023). Substructure-masking methods for GNNs also focus on generating explanations as explicit chemical substructures and structure-property relationships (Wu et al., 2023). More generally, the recent pairing of predictive optimization with transparent feature attribution, through the lens of automated molecular modeling frameworks, has also extended to physicochemical prediction, such as lipophilicity tasks (Mohan & Chang, 2024).

The explanations at the level of the descriptors can also serve as a starting point for formulating hypotheses, as properties can be intentionally varied in the process of molecular design. In actual experimental conditions, these signals may be used to direct specific structural modifications and the resulting measurements could be used to assess if the expected changes in lipophilicity are observed outside the modeled chemical space.

There are some restrictions to the interpretation of these findings. The synthetic endpoint was created to ensure that chemically plausible signals in the structure-property relationships are maintained; high predictive results are therefore not necessarily meaningful in terms of validation against laboratory measurements. Secondly, the train-test split is performed in a random fashion, preserving the distribution, which is not a rigorous scaffold or temporal external validation. Third, descriptors can be correlated, meaning that permutation importance and SHAP importance could share credit between related variables. Future research should then replicate the framework on experimentally measured logP data, use scaffold-based external validation, consider applicability domains and uncertainty, as well as compare descriptor models with fingerprints, graph neural networks and molecular language models. A recent study that combines explainable machine learning and language-based reasoning demonstrates the exciting potential to transform statistical structure-property relationships into human-readable chemical hypotheses (Wellawatte & Schwaller, 2025). Such extensions would define whether or not the transparent relationships found here would hold true in real new chemical space.

5. Conclusion

In this study an interpretable machine-learning framework to trace molecular lipophilicity from chemical structure has been established based on a controlled synthetic data set. Coherent relationships with the synthetic LogP endpoint and the descriptors of molecular polarity and hydrogen-bonding capacity, carbonaceous size, refractivity, flexibility, and related structural properties were found in the statistical analysis. The lipophilicity was found to be decreased by TPSA and descriptors related to hydrogen-bonding, while the descriptors related to carbon-rich and size-related descriptors, in general, contributed in the opposite sense. Comparative regression modeling demonstrated this ability to capture these multivariable relationships in a nonlinear way and the best cross-validated model had a good predictive accuracy on the withheld test subset. Importantly, permutation importance and SHAP analysis translated predictive performance into an interpretable hierarchy of molecular determinants to examine the model in chemical, not just statistical, terms. Results indicate that descriptor-based interpretable machine learning can be used as a transparent link between the molecular structure and the prediction of physicochemical properties. But, the results were proof of concept as the endpoint used was synthetic and the results should not be taken as proof of experiment logP prediction. Future work should be conducted with experimentally measured, chemically diverse data by the same workflow, include external validation/assessment of applicability via scaffolds, and make comparisons between conventional descriptors, fingerprints, and graph-based representations. This validation would determine if the observed structure-lipophilicity relationships extend to more practical situations in medicinal chemistry and if a sensible model, that could be interpreted in terms of structure, could give a good guide for molecular design and prioritisation.

References

1. Aliagas, I., Gobbi, A., Lee, M. L., & Sellers, B. D. (2022). Comparison of logP and logD correction models trained with public and proprietary data sets. *Journal of Computer-Aided Molecular Design*, 36, 253-262. <https://doi.org/10.1007/s10822-022-00450-9>
2. Comesana, A. E., Huntington, T. T., Scown, C. D., Niemeyer, K. E., & Rapp, V. H. (2022). A systematic method for selecting molecular descriptors as features when training models for predicting physiochemical properties. *Fuel*, 321, 123836.
3. Datta, R., Das, D., & Das, S. (2021). Efficient lipophilicity prediction of molecules employing deep-learning models. *Chemometrics and Intelligent Laboratory Systems*, 213, 104309. <https://doi.org/10.1016/j.chemolab.2021.104309>
4. Guha, R., & Velegol, D. (2023). Harnessing Shannon entropy-based descriptors in machine learning models to enhance the prediction accuracy of molecular properties. *Journal of Cheminformatics*, 15, 54.
5. Heberle, H., Zhao, L., Schmidt, S., Wolf, T., & Heinrich, J. (2023). XSMILES: Interactive visualization for molecules, SMILES and XAI attribution scores. *Journal of Cheminformatics*, 15, 2.
6. Jiang, D., Wu, Z., Hsieh, C. Y., Chen, G., Liao, B., Wang, Z., Shen, C., Cao, D., Wu, J., & Hou, T. (2021). Could graph neural networks learn better molecular representation for drug discovery? A comparison study of descriptor-based and graph-based models. *Journal of Cheminformatics*, 13, 12.
7. Jimenez-Luna, J., Grisoni, F., & Schneider, G. (2020). Drug discovery with explainable artificial intelligence. *Nature Machine Intelligence*, 2, 573-584.
8. Kong, Y., Zhao, X., Liu, R., Yang, Z., Yin, H., Zhao, B., Wang, J., Qin, B., & Yan, A. (2022). Integrating concept of pharmacophore with graph neural networks for chemical property prediction and interpretation. *Journal of Cheminformatics*, 14, 52.
9. Lenselink, E. B., & Stouten, P. F. W. (2021). Multitask machine learning models for predicting lipophilicity (logP) in the SAMPL7 challenge. *Journal of Computer-Aided Molecular Design*, 35, 901-909. <https://doi.org/10.1007/s10822-021-00405-6>
10. Li, X., & Fourches, D. (2020). Inductive transfer learning for molecular activity prediction: Next-Gen QSAR Models with MolPMoFiT. *Journal of Cheminformatics*, 12, 27.
11. Mastropietro, A., Pasculli, G., Feldmann, C., Rodriguez-Perez, R., & Bajorath, J. (2022). EdgeSHAPer: Bond-centric Shapley value-based explanation method for graph neural networks. *iScience*, 25(10), 105043.
12. Mohan, B., & Chang, J. (2024). Chemical SuperLearner (ChemSL)-An automated machine learning framework for building physical and chemical properties model. *Chemical Engineering Science*, 294, 120111.
13. Ponzoni, I., Paez Prosper, J. A., & Campillo, N. E. (2023). Explainable artificial intelligence: A taxonomy and guidelines for its application to drug discovery. *WIREs Computational Molecular Science*, 13(6), e1681.
14. Prasad, S., & Brooks, B. R. (2020). A deep learning approach for the blind logP prediction in SAMPL6 challenge. *Journal of Computer-Aided Molecular Design*, 34, 535-542.
15. Ren, G. P., Yin, Y. J., Wu, K. J., & He, Y. (2023). Force field-inspired molecular representation

- learning for property prediction. *Journal of Cheminformatics*, 15, 17.
16. Rodriguez-Perez, R., & Bajorath, J. (2021). Explainable machine learning for property predictions in compound optimization. *Journal of Medicinal Chemistry*, 64(24), 17744-17752.
 17. Ross, J., Belgodere, B., Chenthamarakshan, V., Padhi, I., Mroueh, Y., & Das, P. (2022). Large-scale chemical language representations capture molecular structure and properties. *Nature Machine Intelligence*, 4, 1256-1264.
 18. Stepisnik, T., Skrlj, B., Wicker, J., & Kocev, D. (2021). A comprehensive comparison of molecular feature representations for use in predictive modeling. *Computers in Biology and Medicine*, 130, 104197.
 19. Tang, B., Kramer, S. T., Fang, M., Qiu, Y., Wu, Z., & Xu, D. (2020). A self-attention based message passing neural network for predicting molecular lipophilicity and aqueous solubility. *Journal of Cheminformatics*, 12, 15. <https://doi.org/10.1186/s13321-020-0414-z>
 20. Wang, Y., Wang, J., Cao, Z., & Barati Farimani, A. (2022). Molecular contrastive learning of representations via graph neural networks. *Nature Machine Intelligence*, 4, 279-287.
 21. Wellawatte, G. P., & Schwaller, P. (2025). Human interpretable structure-property relationships in chemistry using explainable machine learning and large language models. *Communications Chemistry*, 8, 11.
 22. Win, Z. M., Cheong, A. M. Y., & Hopkins, W. S. (2023). Using machine learning to predict partition coefficient (Log P) and distribution coefficient (Log D) with molecular descriptors and liquid chromatography retention time. *Journal of Chemical Information and Modeling*, 63(7), 1906-1913.
 23. Withnall, M., Lindelof, E., Engkvist, O., & Chen, H. (2020). Building attention and edge message passing neural networks for bioactivity and physical-chemical property prediction. *Journal of Cheminformatics*, 12, 1.
 24. Wu, Z., Wang, J., Du, H., Jiang, D., Kang, Y., Li, D., Pan, P., Deng, Y., Cao, D., Hsieh, C. Y., & Hou, T. (2023). Chemistry-intuitive explanation of graph neural networks for molecular property prediction with substructure masking. *Nature Communications*, 14, 2585.
 25. Yang, C. I., & Li, Y. P. (2023). Explainable uncertainty quantifications for deep learning-based molecular property prediction. *Journal of Cheminformatics*, 15, 13.