Less is More: XAI-Driven Feature Selection for Efficient Molecular Property Prediction

Jonas Lucas Durão[1], Luís G. Dias[2], and Marcos G. Quiles[1(✉)]

[1] Institute of Science and Technology, Federal University of São Paulo, Av. Cesare Lattes, 1201, São José dos Campos, SP 12247-014, Brazil
{jonas.durao,quiles}@unifesp.br

[2] Chemistry Department, FFCLRP, University of São Paulo, Ribeirão Preto, SP 14040–901, Brazil
lgdias@ffclrp.usp.br

Abstract. Molecular descriptors are commonly employed to build predictive models for molecular properties; however, their usage comes with several challenges. The exponential growth of available descriptors and the complexity of the required algorithms often result in opaque, 'black-box' models that lack interpretability. Traditional feature selection methods focus on statistical optimization, but they often fail to capture the physicochemical drivers and complex interactions central to the chemical domain. In this work, we propose an explainability-driven feature selection framework that uses Shapley Additive exPlanations (SHAP) to identify the minimal necessary subset of descriptors for predicting some molecular properties of the QM9 dataset. This framework serves not only as a dimensionality reduction tool but also as a validation mechanism that ensures the model's logic is consistent with the laws of chemistry. By iteratively reducing the initial set of 1057 Mordred descriptors, we show that feature selection using SHAP values might outperform traditional methods in several scenarios, such as predicting Rotational Constants C and B, Internal Energy at 0 K, and Electronic Spatial Extent. For other molecular properties, the SHAP-guided feature selection delivers similar performance to baseline feature selection methods, such as LASSO, Random Forest, CFS, as well as Pearson and Spearman correlations. Our findings show that the feature space can be compressed by more than 90% with no loss in predictive capability. For some thermodynamic and electronic properties, the SHAP-guided feature reduction has also improved generalization performance due to the elimination of non-informative features. More importantly, the variables surviving this selection process are not simply mathematical conveniences. They map directly onto recognized physicochemical theory, pinpointing the precise electronic and topological mechanisms responsible for molecular behavior.

Keywords: Explainable Artificial Intelligence · XAI · Feature Selection · QM9 dataset · Machine Learning · Molecular Property Prediction

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2027
O. Gervasi et al. (Eds.): ICCSA 2026, LNCS 16767, pp. 204–217, 2027.
https://doi.org/10.1007/978-3-032-30491-9_14

1 Introduction

One of the challenges in machine learning is data dimensionality; over the years, the number of features has increased, and with that came the need for algorithms and tools to address this problem [6]. Therefore, the main objective of feature selection is to reduce the number of input features and lower computational costs while enhancing predictive performance and generalization capabilities of models [15,16].

Traditional feature selection approaches, such as filter, wrapper, and embedded methods [3], often face significant limitations. They are effective in optimizing error metrics, but traditional methods disregard the model's internal logic, failing to reveal the behavior and meaning behind the decisions of the "black box" models [1,2]. Conventional protocols frequently fail to delineate intricate, non-linear interactions across the feature space. Compounding this limitation is the absence of a cohesive metric capable of reconciling raw predictive power with structural transparency [4].

In the Chemistry domain, that requires the understanding of the physical drivers behind a prediction, the lack of interpretation of features selected by traditional feature selection methods makes the final model difficult to understand. This gap can be filled with a shift towards methods that can provide explanations of the results to justify the inclusion of specific descriptors in the input data [5].

In that sense, Explainable Artificial Intelligence (XAI) methods have emerged as powerful tool for feature selection [8]. The integration of XAI based methods into the feature selection phase allows the usage of model explanations to find significant features, and this makes the selection problem more understandable [10]. In this work, we employ SHAP [7] as the core XAI method for feature selection. The choice of SHAP is not only because its a traditional method in the XAI literature, but unlike other methods, SHAP provides a mathematically fair allocation of importance, specifically because of its properties: Efficiency, Symmetry, Null Player, and Linearity [13].

The extensive characterization of molecular topology and physicochemistry that Mordred offers generates a vast feature space (1,613 2D and 213 3D descriptors) populated by redundancy and multicollinearity. Consequently, separating the contribution of each descriptor is a difficult challenge. In this context, because of the SHAP properties, it can act as a critical filter to identify the ideal subset of chemical descriptors necessary for predicting each property. Leveraging these properties allows us to fairly discriminate between important and irrelevant features within the initial feature set.

The application of XAI driven feature selection advances property prediction by providing a different approach to identifying descriptors that not only can improve the final prediction but also show the chemically relevant descriptors for properties. This work bridges the gap between the high accuracy of complex "black-box" models and the necessity for chemically meaningful interpretations of the outputs. By adopting this "Less is More" idea, we ensure an ideal trade-off

between predictive performance, computational efficiency, and the interpretability required for scientific discovery in Chemistry [14].

2 Methodology

2.1 QM9 Dataset and Molecular Representation

For building the models, we used the QM9 dataset [12]. This dataset is widely recognized as a gold standard in the fields of quantum chemistry and machine learning. It consists of approximately 134,000 stable organic molecules, composed of up to nine heavy atoms of four species (C, O, N, F), in addition to hydrogen.

The physicochemical properties of each molecule in QM9 were originally calculated using Density Functional Theory (DFT) at the B3LYP/6-31G(2df,p) level of theory. In this work, all available energetic, electronic, and thermodynamic properties were considered, with the exception of heat capacity (C_v).

2.2 Featurization

The numerical representation of chemical structures was performed using the Mordred library [9]. Unlike approaches based solely on binary fingerprints, Mordred calculates continuous physicochemical descriptors and topological indices based on graph theory. This enhances the model's ability to capture valuable information about the molecules [11].

For this study, an initial vector containing 1057 descriptors was generated for each molecule of QM9. The full 1826 descriptors from the Mordred was not used because some exhibited errors or NaN values when calculating, so they were discarded remaining 1057 descriptors. This set served as the baseline for subsequent feature selection and optimization processes.

2.3 Experimental Setup

To rigorously evaluate the proposed explainability-driven feature selection pipeline, we established a robust validation framework. Prior to model training, a strict preprocessing protocol was applied to the feature space $\mathcal{X}$. Features were standardized to zero mean and unit variance (z-score normalization) to mitigate scale disparity issues inherent in heterogeneous physicochemical descriptors. The dataset was randomly split into training 80%, validation 10%, and test 10% sets using a fixed random seed to ensure reproducibility.

All feature selection methods shared the same underlying predictive network. The model employed for property prediction is a Multilayer Perceptron (MLP). To ensure a correct assessment of feature importance, regardless of the target property's complexity, we treated the model architecture as a hyperparameter search problem rather than a fixed topology. We employed the Tree-structured Parzen Estimator (TPE) algorithm, via the Optuna framework, to optimize the architecture and training hyperparameters for each of the 15 properties independently. The search space encompassed:

1. **Architecture:** Depth (1 to 4 hidden layers) and width (64 to 1024 neurons per layer).
2. **Optimization:** Learning rate (10^{-5} to 10^{-1}), batch size (128 to 1024), and optimizer selection (Adam, RMSprop, SGD, AdamW).
3. **Regularization:** Dropout rates (0.0 to 0.75) and weight decay (L_2 regularization).
4. **Objective Function:** A categorical choice between L_1 (MAE), MSE, and Huber (Smooth L_1) loss functions.

Early stopping with a patience of 10 epochs was utilized to prevent overfitting during the optimization trials. All experiments were conducted on a workstation equipped with two 12-core Intel Xeon Silver 4214R processors, 256 GB of RAM, and 2 NVidia RTX A6000 GPUs, each with 48 GB of memory.

2.4 Feature Selection and Degradation Protocol

We implemented an iterative feature degradation protocol Fig. 1 to assess the explanation of the identified feature subsets. The core utilization of this protocol is to determine the minimal subset of features $S \subset \mathcal{D}$ required to maintain predictive performance comparable to the full feature set, in which $\mathcal{D}$ represents the full feature set. The iterative pruning process is defined by:

$$S_k = S_{k-1} \setminus \{d_j \in S_{k-1} \mid \Omega_j^{(k)} < \tau_k\} \tag{1}$$

where S_k is the feature subset at iteration k, and $\Omega_j^{(k)}$ represents the importance score of descriptor d_j calculated by the chosen method. Following each pruning step, the model f is retrained on the reduced subspace $\mathcal{X}_{S_k}$ until the termination criterion is met.

The degradation process was terminated at 20 features for all methods based on the importance score, features ranked below the 20th position exhibited importance values less than 1%, of the top-ranked feature, indicating negligible contribution to model predictions.

XAI-Driven Selection. We used SHAP global explanations from the toolbox ChemXAI. For a given trained MLP model f, the importance of feature j is quantified as the mean absolute SHAP value across the training cohort:

$$I_j^{SHAP} = \frac{1}{N} \sum_{i=1}^{N} |\phi_j(f, x^{(i)})| \tag{2}$$

where ϕ_j represents the contribution of feature j to the prediction of sample $x^{(i)}$. In each iterative step, we keep the top 90% of the features with the highest global importance.

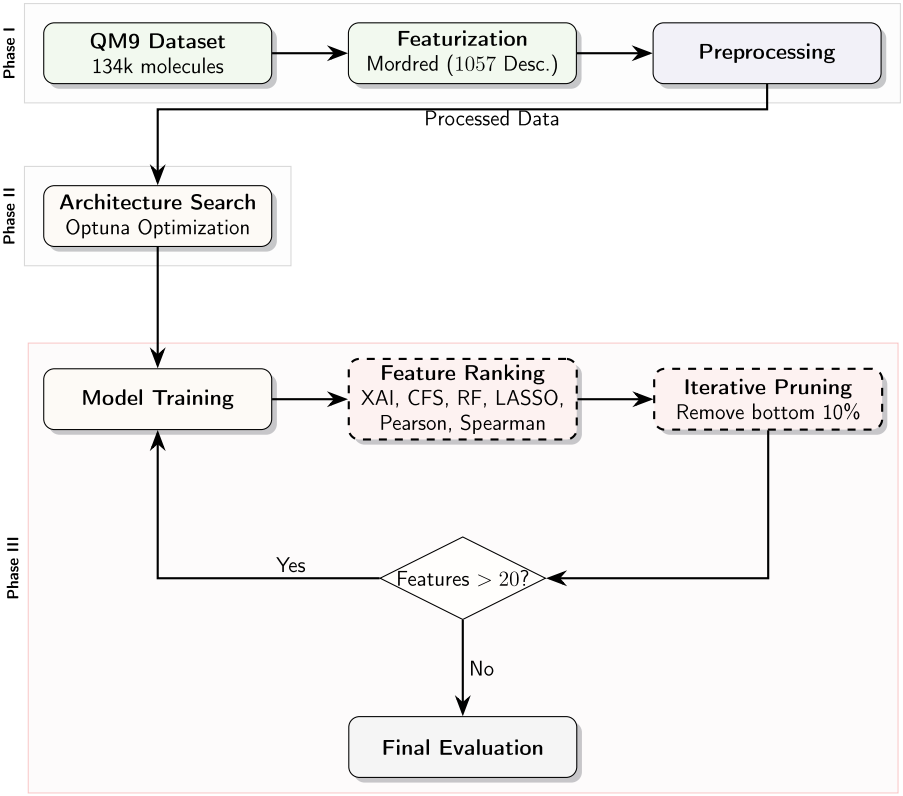


Fig. 1. Iterative Feature Pruning and Model Re-training Workflow.

Baseline Methods. To validate the efficacy of the SHAP-based selection, we benchmark against five well-established feature selection baselines representing filter and embedded methods:

LASSO (Least Absolute Shrinkage and Selection Operator): An embedded method that enforces sparsity through L1 regularization. To ensure computational efficiency in the high-dimensional Mordred space, we utilized a fixed regularization parameter $\alpha = 0.001$ and a maximum of 2,000 iterations. Features are ranked based on the absolute values of their coefficients.

Random Forest Importance (RF): A non-linear embedded method using an ensemble of decision trees. We utilized 50 estimators with a maximum depth of 10 to capture non-linear interactions while maintaining memory safety. Feature importance is computed based on the Mean Decrease in Impurity (MDI).

Pearson Correlation: A linear filter method that measures the strength and direction of the linear relationship between each descriptor and the target property. This serves as a baseline for detecting strictly linear dependencies.

Spearman Rank Correlation: A non-parametric filter method that assesses monotonic relationships. It is implemented by calculating the Pearson correlation

on the rank-transformed variables, allowing for the detection of non-linear but monotonic dependencies.

CFS (Correlation-based Feature Selection): Functioning as a proxy for CFS, this method utilizes the F-statistic via univariate linear regression (f_regression). It ranks features by evaluating the individual effect of each regressor on the target property.

3 Results and Discussion

For the results presented below, we evaluate the global effectiveness of the XAI-based feature selection methodology across all molecular properties; then, we compare our approach against traditional baseline methods and provide detailed case studies demonstrating the chemical interpretability of the selected features.

3.1 Global Reduction Analysis

With feature degradation using SHAP as the selection method, 11 of the 14 properties (Fig. 2) exhibited improved performance compared to the full feature set of 1057 descriptors. The applied methodology was effective in enhancing model performance by reducing test error, and it successfully eliminated redundant features that added no predictive information while increasing model complexity.

The results depicted in Fig. 2 reveal a heterogeneous response of molecular properties to the XAI feature selection. This result shows a clear improvement for thermodynamic properties that suggests a presence of non-essential features for these particular targets. Acting as a filter, SHAP distills the descriptors into an optimal subset that actively isolates the primary energetic drivers. In contrast, electronic properties such as Dipole Moment (μ) and orbital energies ϵHOMO and ϵLUMO showed no improvement. This indicates a higher intrinsic dimensionality for electronic targets, where a more diverse set of descriptors is needed to account for the variability of the electron density in the dataset, making them less susceptible to performance gains via aggressive feature pruning.

For the bottom 6 properties in Table 2, the method achieved more than a 50% reduction over the original descriptor set without sacrificing predictive capacity, illustrating that the degradation process can act as a denoising step when guided by SHAP. When the loss ratio of first degradation to current degradation stays above unity during feature removal, the XAI-based selection exceeds the performance of the initial full-feature baseline. Figure 3a illustrates this process for the property Rotational Constant C. These results confirm that the proposed approach operates as an active, targeted filter, distinctly moving beyond blind statistical dimensionality reduction.

3.2 Evolution of Test Loss

The optimized models achieved low test losses across all properties, demonstrating robust generalization. The evolution of test loss with feature reduction, however, depends strongly on the nature of the target property (Table 1, Table 2).

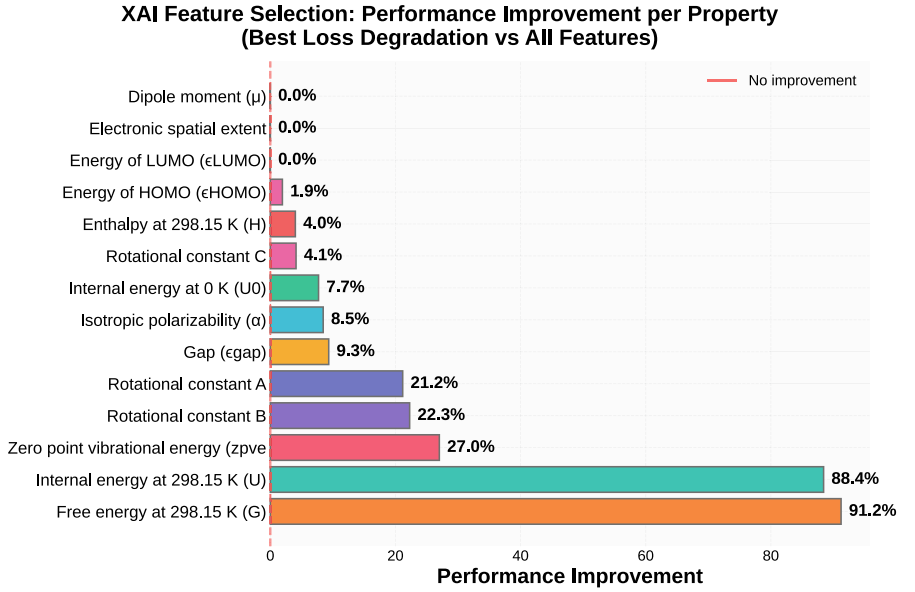
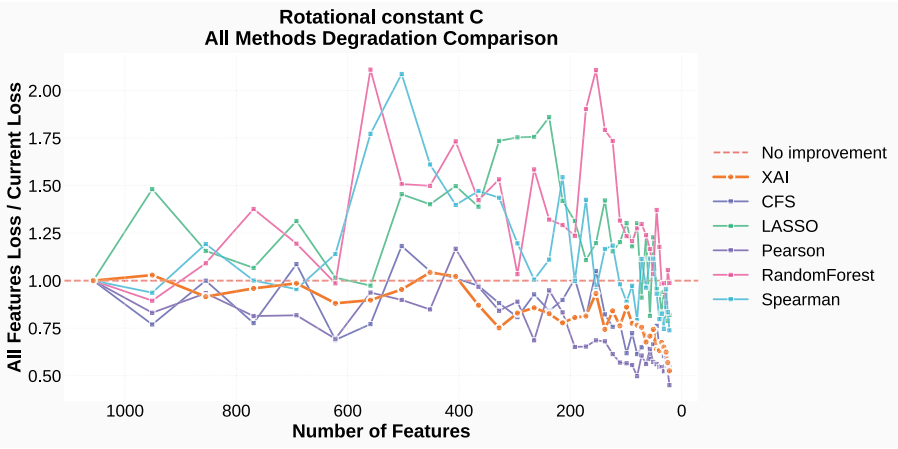


Fig. 2. A comparative performance gains of XAI-driven feature selection across all QM9 properties. The chart illustrates the improvement when moving from the full feature set (1057 descriptors) to the optimal feature subset identified during the iterative degradation protocol.

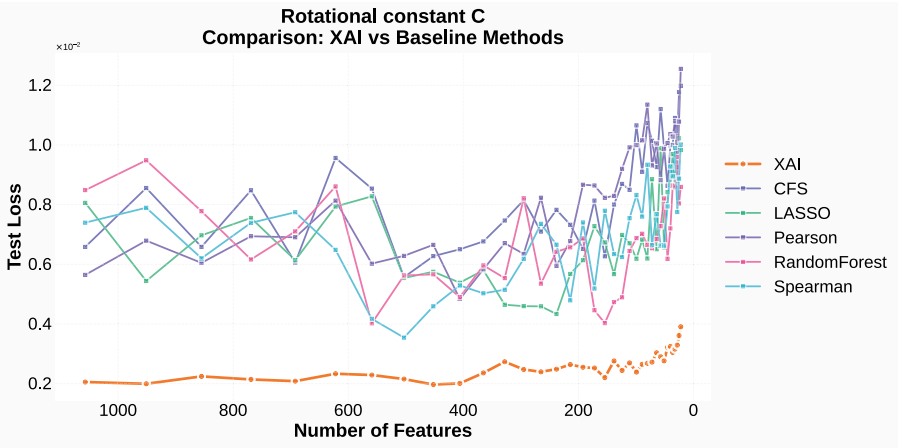
Electronic and vibrational properties (e.g., HOMO-LUMO gap, zero-point vibrational energy) exhibit a clear dependence on descriptor count; the test loss tends to rise as features are pruned, and a larger subset is needed to capture their behavior.

In contrast, thermodynamic and rotational properties reach an elbow point in the end of the degradation process. For these, the test loss stabilizes or oscillates at a low level even under drastic dimensionality reduction, often requiring fewer than 30 features to achieve the best performance. This suggests that thermodynamic and rotational behavior can be captured by a compact set of descriptors. The XAI approach showed that it can offer a favorable trade-off for vibrational and structural targets by retaining a larger subset than some baselines while prioritizing chemically meaningful descriptors and maintaining predictive accuracy at the elbow of the curve. Figure 3b illustrates this behavior for Rotational Constant C, where XAI attains the lowest test loss among all methods and has a more stable oscillation as the feature set is reduced.

In the detailed analysis of the Fig. 3, the superiority of the XAI approach is observed at two levels. In plot Fig. 3a, the XAI curve consistently remains above the 'No improvement' baseline, reaching efficiency peaks even under severe dimensionality reduction. Most notably, plot Fig. 3b demonstrates that the test losses for the baselines increase drastically upon crossing the 200 features threshold; the SHAP based selection maintains a low and stable test error, increasing



(a)



(b)

Fig. 3. Comparative analysis of feature selection impact on the Rotational Constant C prediction. (a) Displays the relative performance expressed as the test loss ratio ($Loss_{full}/Loss_{reduced}$) as features are iteratively removed; (b) Illustrates the absolute performance evolution of the Test Loss across the degradation process for XAI and baseline methods. The shaded areas represent the standard deviation across multiple experimental runs.

at a lower rate than the baselines. This indicates that the algorithm identified geometrically fundamental descriptors that purely statistical methods discarded prematurely.

Table 1. Comparison of Minimum Test Loss (Mean ± Std. Dev.).

Sym.	XAI	CFS	LASSO	Pearson	RF	Spearman
μ [$\times 10^{-1}$]	1.178 ± 0.003	1.164 ± 0.029	$\mathbf{1.146 \pm 0.013}$	1.151 ± 0.005	1.153 ± 0.010	1.155 ± 0.015
R_e [$\times 10^{-2}$]	$\mathbf{1.19 \pm 0.02}$	1.20 ± 0.01	1.20 ± 0.05	1.20 ± 0.03	1.21 ± 0.03	1.21 ± 0.05
ϵ_H [$\times 10^{-2}$]	3.77 ± 0.12	3.51 ± 0.05	3.63 ± 0.06	3.62 ± 0.19	$\mathbf{3.50 \pm 0.02}$	3.63 ± 0.02
ϵ_L [$\times 10^{-3}$]	9.93 ± 0.31	9.83 ± 0.11	$\mathbf{9.60 \pm 0.24}$	9.70 ± 0.07	9.63 ± 0.05	9.78 ± 0.14
H [$\times 10^{-5}$]	5.63 ± 0.21	0.99 ± 0.06	3.45 ± 0.59	1.09 ± 0.26	1.78 ± 0.51	$\mathbf{0.64 \pm 0.04}$
G [$\times 10^{-5}$]	2.65 ± 0.63	$\mathbf{0.51 \pm 0.11}$	2.00 ± 0.85	1.00 ± 0.73	2.02 ± 0.44	0.93 ± 0.87
ϵ_{gap} [$\times 10^{-2}$]	1.60 ± 0.11	1.61 ± 0.13	1.60 ± 0.08	1.65 ± 0.12	$\mathbf{1.54 \pm 0.05}$	1.65 ± 0.13
U_0 [$\times 10^{-4}$]	$\mathbf{4.86 \pm 0.53}$	5.06 ± 0.44	5.16 ± 0.12	5.81 ± 0.29	5.31 ± 0.13	5.73 ± 0.32
U [$\times 10^{-5}$]	2.87 ± 1.26	3.37 ± 0.25	$\mathbf{0.71 \pm 0.86}$	4.28 ± 1.85	2.66 ± 1.53	3.76 ± 0.51
α [$\times 10^{-3}$]	5.46 ± 0.37	5.66 ± 0.49	$\mathbf{5.32 \pm 0.21}$	5.85 ± 0.19	5.71 ± 0.14	5.61 ± 0.40
A [$\times 10^{-7}$]	9.07 ± 0.03	3.20 ± 0.05	2.90 ± 0.03	2.97 ± 0.02	3.03 ± 0.02	$\mathbf{2.87 \pm 0.01}$
B [$\times 10^{-3}$]	$\mathbf{3.20 \pm 0.21}$	4.82 ± 0.81	3.93 ± 0.75	4.44 ± 0.29	4.22 ± 1.15	5.29 ± 0.64
C [$\times 10^{-3}$]	$\mathbf{1.97 \pm 0.09}$	5.57 ± 2.03	4.33 ± 1.27	4.84 ± 0.95	4.02 ± 0.77	3.54 ± 0.51
E_{zpve} [$\times 10^{-4}$]	3.3 ± 0.2	3.2 ± 0.2	3.2 ± 0.0	$\mathbf{2.9 \pm 0.1}$	3.2 ± 0.3	3.0 ± 0.2

Properties: μ: Dipole moment; R_e: Electronic spatial extent; ϵ_H: Energy of HOMO; ϵ_L: Energy of LUMO; H: Enthalpy (298.15 K); G: Free energy (298.15 K); ϵ_{gap}: Gap energy; U_0: Internal energy (0 K); U: Internal energy (298.15 K); α: Isotropic polarizability; A, B, C: Rotational constants; E_{zpve}: Zero point vib. energy. **Note:** Multipliers in brackets (e.g., [$\times 10^{-2}$]) apply to both mean and std. dev.

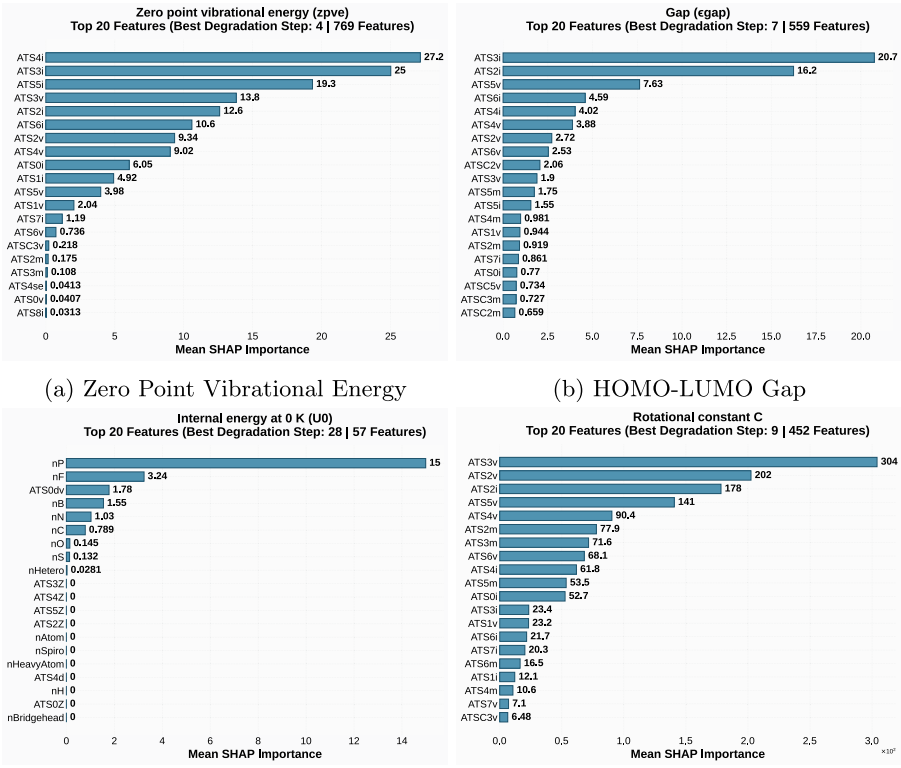
3.3 Case Study

The SHAP values across Fig. 4 reveal a big disparities in sensitivity scales among the property classes. HOMO-LUMO Gap and ZPVE, exhibit importance scores between 20 and 30. In comparison, the Rotational Constant C values are on a scale roughly ten times larger, with the top descriptor reaching a peak value of 304.

This order of magnitude difference highlights the extreme sensitivity of molecular inertia to specific topological volume distributions. This means that, for rotational constants, the volume of the molecule, $ATSXv$, reflects the prediction more directly than other properties, which require more descriptors to explain their behavior. A similar trend appears for Internal Energy at 0 K, where the phosphorus count, nP, is nearly five times as influential as fluorine. These steep gradients in importance justify the degradation of the features. Since a small set of high valuable descriptors captures the target, the model can undergo aggressive feature pruning without significant accuracy loss.

The top 20 SHAP rankings (Fig. 4) also reveal how the selected descriptors align with physicochemical intuition. For electronic and vibrational targets (Fig. 4a, Fig. 4b), the model attributes high importance to the Autocorrelation of Topological Structure (ATS) descriptors, weighted by ionization potential ($ATSXi$) and van der Waals volume ($ATSXv$). The prominence of $ATS3i$, $ATS2i$, and $ATS4i$ is consistent with the sensitivity of gap and zero-point energy to charge distribution and molecular size at short topological distances. Thermo-

dynamic properties are driven by descriptors that reflect elemental composition and electronic distribution. For Internal Energy at 0 K (Fig. 4c), heteroatom counts (nP, nF, nB, nN) rank among the most important features, in line with the dependence of formation energy on atomic identity and bond contributions.



(a) Zero Point Vibrational Energy

(b) HOMO-LUMO Gap

(c) Internal Energy at 0 K (U_0)

(d) Rotational Constant C

Fig. 4. Global feature importance ranking based on Mean SHAP values for four selected QM9 properties. Each panel illustrates the top 20 descriptors identified at the best degradation step: (a) Zero Point Vibrational Energy, (b) HOMO-LUMO Gap, (c) Internal Energy at 0 K, and (d) Rotational Constant C. The x-axis represents the average absolute contribution of each feature to the model's output across the test set, highlighting the dominance of specific topological and electronic indices in governing these molecular properties.

When XAI does not achieve the lowest test loss, the chosen features remain chemically meaningful. For Dipole moment, the baselines such as Pearson or LASSO achieved better test errors, and the dominant SHAP features (e.g., $ATS2v$, $ATS2i$) still reflect charge distribution and size. Thus, the model identifies interpretable drivers despite of minor differences in predictive performance.

In contrast, for Rotational Constant C, XAI attains the lowest test loss (Table 1) and the descriptors (Fig. 4d) agree with the geometric and inertial drivers of rotational constants. Figure 3a shows that the loss ratio remains stable and above 1.0 for the XAI curve as features are removed, confirming that SHAP based feature selection acts as denoising rather than arbitrary removal.

3.4 Comparisons to Feature Selection Methods

Comparison with traditional baselines (Table 1, Table 2) highlights both the competitiveness and the distinctive value of the XAI approach. For Internal Energy at 0 K (U0), XAI achieved the best loss and required only 57 features compared to 406 for LASSO. This example illustrates that XAI offers a viable alternative to LASSO, with the added benefit that the retained descriptors align with chemical intuition (Fig. 4c).

Table 2. Comparison of the Number of Selected Features for the Lowest Test Loss across XAI and Baseline Methods for QM9 Properties.

Property	XAI	CFS	LASSO	Pearson	RF	Spearman
μ	1057	769	1057	855	622	855
R_e	1057	769	769	951	692	1057
ϵ_H	855	951	692	1057	214	622
ϵ_L	1057	951	1057	855	951	951
H	951	22	22	28	25	22
G	64	22	25	28	22	25
ϵ_{gap}	559	622	192	452	214	559
U_0	57	51	406	1057	855	64
U	45	22	25	22	28	28
α	951	855	265	1057	855	951
A	192	855	769	769	40	28
B	99	452	124	192	214	855
C	452	406	111	214	406	503
E_{zpve}	769	559	951	855	1057	769

Legend: RF: Random Forest. μ: Dipole moment; R_e: Electronic spatial extent; ϵ_H: Energy of HOMO; ϵ_L: Energy of LUMO; H: Enthalpy (298.15 K); G: Free energy (298.15 K); ϵ_{gap}: Gap energy; U_0: Internal energy (0 K); U: Internal energy (298.15 K); α: Isotropic polarizability; A, B, C: Rotational constants; E_{zpve}: Zero point vib. energy.

For Internal Energy at 298.15 K (U), XAI achieved the third highest performance gain with one of the smallest feature sets. The network complexity

dropped from 1,253,057 parameters (1057 features) to 346,305 parameters (45 features), a 72.4% reduction without compromising accuracy. For larger architectures, such savings would translate into substantial gains in training time and memory. Moreover, even in cases where another method reports a marginally lower test loss, XAI provides transparency over which features drive the prediction, shifting the focus from purely statistical optimization to chemically grounded, understandable models.

The descriptors selected by the different methods reveal a clear distinction between the baselines and XAI. LASSO, CSF, and Random Forest frequently rely on generic topological indices and constitutional counts (e.g., nAtom, SpAbs_A) for most of the properties, showing that it does not search for chemically meaningful features that describe the property. For the XAI selected features, SHAP gives more importance and highlights Moreau-Broto autocorrelations weighted by ionization potential and van der Waals volumes (ATS4i, ATSC3v), which accurately reflect spatial electron distribution for electronic properties. Similarly, for thermodynamic properties such as Enthalpy (H), XAI selects descriptors related to mass distribution and internal connectivity (ATS3m, ATS5i) rather than purely statistical correlations.

4 Conclusions

We have demonstrated that explainability driven feature selection based on SHAP attributions substantially reduced the original set of Mordred descriptors for QM9 molecular property prediction. This approach maintains or improves predictive performance, with 11 of 14 targets benefiting from SHAP-based reduction compared to the 1057-feature initially. Six properties achieved a reduction of over 50% while simultaneously lowering test error. Thermodynamic and rotational properties reached optimal performance with fewer than 30 descriptors, whereas electronic and vibrational targets required larger subsets. This heterogeneity reflects the underlying physics, as global descriptors are sufficient for thermodynamic and rotational behavior, and it demonstrates that energy gaps and zero-point energy depend on finer electronic and structural details.

The SHAP based approach matches traditional methods like LASSO, Random Forest importance, CFS, Pearson Correlation, and Spearman Ranking on several properties, including Rotational Constants C and B, Electronic Spatial Extent, and Internal Energy at 0 K. It remains competitive for Internal Energy at 298.15 K using significantly fewer features. The retained subsets including ATS type autocorrelations for electronic targets, heteroatom counts for thermodynamic energies, and geometry related descriptors for rotational constants align with physicochemical intuition to provide understandable models. Even when alternative methods provide marginally lower test loss, the interpretability and trust of the SHAP selected set offer a superior trade-off.

These results establish XAI as a functional tool for feature selection. Integrating iterative degradation with global SHAP importance offers compact, chemically meaningful descriptor sets that reduce computational costs without sacrificing accuracy. Future work may extend this framework to other descriptor

families, model architectures, and larger datasets where efficiency and interpretability are critical.

5 Data and Software Availability

The QM9 dataset is available at https://quantum-machine.org/datasets/. The source code implementing the XAI-driven feature selection framework and the associated validation pipelines will be made publicly available in a GitHub repository upon acceptance of this manuscript for publication.

Acknowledgments. The authors gratefully acknowledge support from FAPESP (São Paulo Research Foundation) and Shell, projects Numbers 2017/11631-2, 2022/09285 – 7, and 2025/01862 – 3, the Brazilian National Council for Scientific and Technological Development (CNPq), grant 310377/2025-0, and the strategic importance of the support given by ANP (Brazil's National Oil, Natural Gas and Biofuels Agency) through the R&D levy regulation.

References

1. Adadi, A., Berrada, M.: Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). IEEE Access **6**, 52138–52160 (2018). https://doi.org/10.1109/ACCESS.2018.2870052
2. Arrieta, A.B., et al.: Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Inf. Fusion **58**, 82–115 (2020). https://doi.org/10.1016/j.inffus.2019.12.012
3. Chandrashekar, G., Sahin, F.: A survey on feature selection methods. Comput. Electr. Eng. **40**(1), 16–28 (2014). https://doi.org/10.1016/j.compeleceng.2013.11.024
4. Fryer, D., Strümke, I., Nguyen, H.: Shapley values for feature selection: the good, the bad, and the axioms. IEEE Access **9**, 144352–144360 (2021). https://doi.org/10.1109/ACCESS.2021.3119110
5. Gomes-Esteves, J., Mendes, G., Faria, P., Vale, Z.: Leveraging explainable artificial intelligence in solar photovoltaic mappings: model explanations and feature selection. Energies **18**(5), 1282 (2025). https://doi.org/10.3390/en18051282
6. Guyon, I., Elisseeff, A.: An introduction to variable and feature selection. J. Mach. Learn. Res. **3**, 1157–1182 (2003)
7. Lundberg, S.M., Lee, S.I.: A unified approach to interpreting model predictions. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. NIPS'17, Red Hook, NY, USA, pp. 4768–4777. Curran Associates Inc. (2017)
8. Marcílio, W.E., Eler, D.M.: From explanation to feature selection: an approach based on local explanations. Comput. Phys. Commun. **256**, 107455 (2020). https://doi.org/10.1016/j.cpc.2020.107455
9. Moriwaki, H., Tian, Y.S., Kawashita, N., Takagi, T.: Mordred: a molecular descriptor calculator. J. Cheminform. **10**(4) (2018). https://doi.org/10.1186/s13321-018-0258-y

10. Munawar, M.A., et al.: Optimizing solar radiation prediction with ANN and explainable AI-based feature selection. Axioms **13**(7), 263 (2024). https://doi.org/10.3390/axioms13070263
11. Pinheiro, G.A., Mucelini, J., Soares, M.D., Prati, R.C., Da Silva, J.L.F., Quiles, M.G.: Machine learning prediction of nine molecular properties based on the smiles representation of the qm9 quantum-chemistry dataset. J. Phys. Chem. A **124**(47), 9854–9866 (2020). https://doi.org/10.1021/acs.jpca.0c05969
12. Ramakrishnan, R., Dral, P.O., Rupp, M., von Lilienfeld, O.A.: Quantum chemistry structures and properties of 134 kilo molecules. Sci. Data **1**, 140022 (2014). https://doi.org/10.1038/sdata.2014.22
13. Rozemberczki, B., et al.: The shapley value in machine learning. In: Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence (IJCAI-22), pp. 5572–5579 (2022). https://doi.org/10.24963/ijcai.2022/782
14. Schreiber, M., Hochberg, K.: Explainable machine learning and predictive statistics for sustainable photovoltaic power prediction on areal meteorological variables. Appl. Sci. **15**(14), 8005 (2025). https://doi.org/10.3390/app15148005
15. Zacharias, J., Zahn, M., Chen, J., Hinz, O.: Designing a feature selection method based on explainable artificial intelligence. Electron. Mark. **32**(4), 2175–2204 (2022). https://doi.org/10.1007/s12525-022-00608-1
16. van Zyl, C., Ye, X., Naidoo, R.: Harnessing explainable artificial intelligence for feature selection in time series energy forecasting: a comparative analysis of grad-cam and shap. Appl. Energy **353**, 122079 (2024). https://doi.org/10.1016/j.apenergy.2023.122079