Systems & Control Transactions

Research Article - Peer Reviewed Conference Proceeding
ESCAPE 35 - European Symposium on Computer Aided Process Engineering
Ghent, Belgium. 6-9 July 2025
Jan F.M. Van Impe, Grégoire Léonard, Satyajeet S. Bhonsale,
Monika E. Polańska, Filip Logist (Eds.)

PSE PRESS

A White-Box AI Framework for Interpretable Global Warming Potential Prediction

Jaewook Lee[a], Ethan Errington[a] and Miao Guo[a]*

[a] Department of Engineering, King's College London, London, WC2R 2LS, United Kingdom
* Corresponding Author: miao.guo@kcl.ac.uk

ABSTRACT

Accurate yet interpretable prediction of Global Warming Potential (GWP) is essential for the sustainable design of chemical products and processes. However, existing studies that utilize molecular structure and physicochemical properties for GWP prediction often suffer from low interpretability, relying on black-box models that obscure the underlying relationships between molecular descriptors and environmental impact. To address this limitation, this study employs a Kolmogorov–Arnold Network (KAN) to derive symbolic equations that establish explicit relationships between molecular properties and GWP. By extracting interpretable mathematical expressions, our approach provides a transparent foundation for decision-making in chemical processes and reaction development. Our comparative analysis of machine learning models—including Random Forest, XGBoost, Deep Neural Networks (DNN), and KAN—reveals that Mordred descriptors outperform MACCS keys in GWP prediction, emphasizing the importance of physicochemical properties. The proposed KAN model achieves predictive accuracy comparable to conventional deep learning methods while maintaining interpretability, facilitating data-driven and transparent sustainability assessments in the chemical industry.

Keywords: Global Warming Potential (GWP), Life Cycle Assessment (LCA), Explainable Artificial Intelligence (XAI), Kolmogorov–Arnold Network (KAN), Environmental Impact Prediction

INTRODUCTION

The chemicals industry, as one of the largest global manufacturing sectors, significantly contributes to global greenhouse gas (GHG) emissions.[1] Major research efforts have been made to design new solutions that decarbonize the way chemical products are made.[2] This requires generating reliable yet transparent predictions of the environmental profiles of chemical products. However, a great deal of information is missing – hindering the extent to which fully informed decision-making can be employed in practice.

Life Cycle Assessment (LCA) provides a standardized framework to evaluate environmental impacts, with Global Warming Potential (GWP) being a key metric. While significant efforts have been made to develop LCA databases for retrospective analysis, they remain incomplete in coverage. The diversity of chemical products, the time-intensive nature of LCA model development, and the variability in available data all contribute to this limitation. Consequently, many chemical products and processes are inadequately represented in current databases, limiting the scope of informed decision-making.

To address these challenges, data-driven predictive models have emerged as an efficient alternative to traditional LCA modelling. In particular, GWP prediction based on molecular structure and physicochemical properties has gained significant attention due to the high availability and reliability of such features.[3-5] These models leverage existing LCA data to identify patterns and estimate GWP values for compounds that lack direct assessment. However, most existing predictive models rely on black-box methodologies, such as deep neural network (DNN), which suffer from low interpretability.

Several studies have attempted to predict LCA impact categories using machine learning models. For example, Zhu et al (2020). developed a DNN-based model capable of predicting multiple LCA endpoints, achieving high accuracy for some categories but significantly lower performance for ecosystem-related metrics.[5] Similarly, Song et al. achieved strong predictive accuracy for Eco-

indicator 99 (Coefficient of Determination (R^2) up to 0.87) but reported poor GWP prediction accuracy (only 0.48).[3] These results underscore the need for improved modeling approaches specifically designed for GWP prediction.

A critical limitation of existing GWP predictive models is their reliance on increasingly complex machine learning architectures, such as DNN, to enhance predictive performance. While these models improve accuracy, their complexity significantly reduces interpretability, making it difficult to understand the underlying relationships between input features and GWP predictions. Many of these models function as black boxes, limiting their practical applicability in scientific and regulatory contexts. Although explainable AI (XAI) techniques, such as shapley additive explanations (SHAP), have been applied to provide post-hoc explanations, these methods do not fundamentally improve the transparency of the model itself.[6-9] Instead, they serve as an external layer of analysis rather than integrating interpretability into the core predictive framework, highlighting the need for inherently interpretable modeling approaches.

To address the limitations of black-box models, there is a growing need for white-box modeling approaches that provide both accuracy and transparency. Kolmogorov–Arnold Network (KAN) has recently emerged as a promising alternative to DNNs by offering a compact and interpretable model architecture.[10, 11] Unlike traditional neural networks, KAN can extract generalized equations from data, allowing predictions to be directly linked to chemical properties. This feature makes KAN particularly suitable for applications requiring both predictive power and interpretability, such as GWP estimation.

This study evaluates the effectiveness of KAN in developing an interpretable model for GWP prediction. By leveraging molecular descriptors and KAN's symbolic representations, we provide a transparent alternative to existing machine learning models. This approach enhances interpretability in data-driven environmental assessments, supporting informed decision-making and sustainable chemical design.

METHODS

Our methodology for GWP prediction follows a structured approach focused on feature engineering, model benchmarking, and white-box modelling.

Data Collection and Preprocessing

The dataset used in this study is sourced from the Ecoinvent v3.8 database, a widely recognized resource for environmental impact assessment. Our dataset consists of 1,689 GWP values spanning 487 different chemicals, ensuring a diverse and representative chemical space for model training. To mitigate data skewness, we applied log transformation, which reduced the effect of extreme GWP values and improved model stability.

Feature Engineering

To develop a robust predictive framework, we identified and utilized optimal molecular descriptors that capture key structural and physicochemical properties of chemical compounds. Specifically, we employed MACCS keys, a set of 166 binary features representing fundamental molecular functional groups.[12] And Mordred descriptors, which provide a more comprehensive representation of molecular structure and physicochemical properties.[13] These descriptors capture essential structural and chemical characteristics, allowing for accurate GWP predictions based on molecular information.

Model Development

Our study evaluates multiple machine learning models, including random forest[14] (RF), extreme gradient boosting[15] (XGBoost), DNN[16], and KAN[10, 11], to compare their performance in GWP prediction. Model performance was assessed using R^2 to determine the most effective approach for GWP estimation. Among these, KAN was selected as the primary model to investigate its potential to balance predictive accuracy and interpretability by extracting symbolic equations that describe the relationship between molecular descriptors and GWP.

All computational experiments were conducted using Python (version 3.12) in a macOS environment. Random Forest (RF) and XGBoost models were implemented using the scikit-learn library.[17] DNN and KAN models were developed using the PyTorch framework.[18]

White-Box Modelling

To enhance transparency, we leveraged KAN's unique architecture to derive explicit mathematical expressions that describe the relationship between molecular descriptors and GWP. We also established a comparative baseline by developing interpretable models using linear regression and RF, both of which are commonly employed for their explainability. Unlike traditional black-box models, this approach allows predictions to be traced back to underlying chemical properties, improving interpretability and decision-making.

By focusing on molecular descriptors, our approach aims to establish a transparent, interpretable, and reproducible methodology for GWP estimation. This ensures that predictions are solely driven by intrinsic chemical properties, allowing for a clear understanding of molecular contributions to GWP.

RESULTS AND DISCUSSION

Learning Curve Analysis

Learning curve analysis was performed to

determine the optimal data volume for effective training. The dataset was partitioned into subsets based on ascending GWP values, and model performance was evaluated by incrementally increasing the training dataset size. XGBoost was chosen for evaluation due to its strong predictive performance and computational efficiency. As shown in **Fig. 1**, the highest R^2 score was achieved using 0.874 of the available data, after which performance declined due to increased data variability. When using only the data from Fold 1 or 2, negative R^2 values were observed. Although the R^2 score typically ranges from 0 to 1, it can take negative values when the model performs worse than simply predicting the mean, indicating poor generalisation. To ensure optimal predictive performance while maintaining model robustness, only data with GWP values below 5026.3 kg CO_2-eq, corresponding to the upper limit of Fold 8, were included in the training dataset. This threshold was chosen based on the observed sharp decline in model accuracy beyond this point, thereby improving generalizability and reducing the risk of overfitting to extreme values

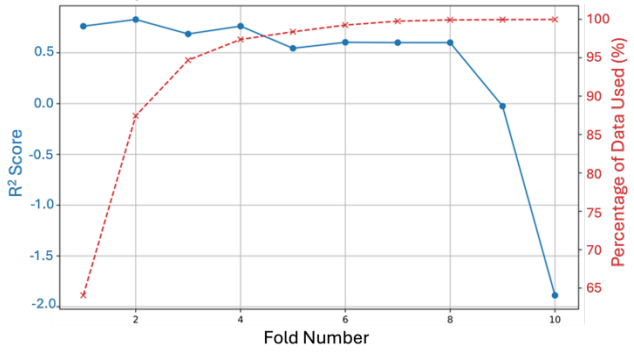

Fig. 1: Learning curve analysis through model performance evaluation across different fold numbers.

Feature Set and Model Performance

We compared the performance of four models—RF, XGBoost, DNN, and KAN—using MACCS keys and Mordred descriptors. **Table 1** shows the R^2 prediction results for each model with different features. Mordred descriptors consistently outperformed MACCS keys across all models, with DNN achieving the highest accuracy. This suggests that physicochemical properties captured by Mordred descriptors play a more significant role in GWP prediction than structural fingerprints alone.

Table 1: Performance comparison of different models and features for GWP prediction

	MACCS keys	Mordred
RF	0.54	0.55
XGBoost	0.64	0.68
DNN	0.60	0.71
KAN	0.57	0.66

The results highlight the necessity of incorporating comprehensive physicochemical descriptors for enhanced model performance. The KAN model demonstrated competitive accuracy while maintaining interpretability, making it a viable alternative to DNNs.

XAI Analysis for Key Feature Identification

To identify the primary factors contributing to GWP prediction, we conducted a post-hoc analysis using the XAI algorithm, SHAP. SHAP quantifies the contribution of each feature to the model's predictions, offering a theoretically grounded approach to explainability in machine learning models.

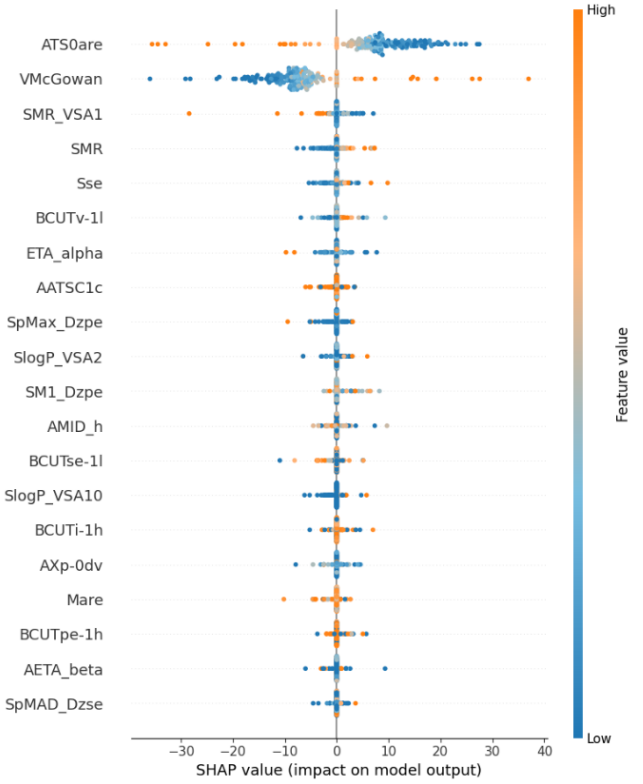

Fig. 2: SHAP value distribution for key chemical descriptors, showing their impact on model output.

For this analysis, we examined the DNN model, which achieved the highest predictive accuracy using chemical descriptors, to gain insights into the key molecular properties influencing GWP. As shown in **Fig. 2**, the analysis identified ATS0are and VMcGowan as the most influential descriptors, both exhibiting a strong linear relationship with model outputs. ATS0are, a Mordred descriptor representing the sum of squared atomic electronegativity properties (specifically, Allred-Rochow electronegativity), showed an inverse correlation with GWP. Lower ATS0are values indicate lower molecular electronegativity, which often correlates with increased atmospheric stability, reduced degradation via photolysis, and prolonged greenhouse gas effects. VMcGowan, which quantifies molecular volume based on McGowan's

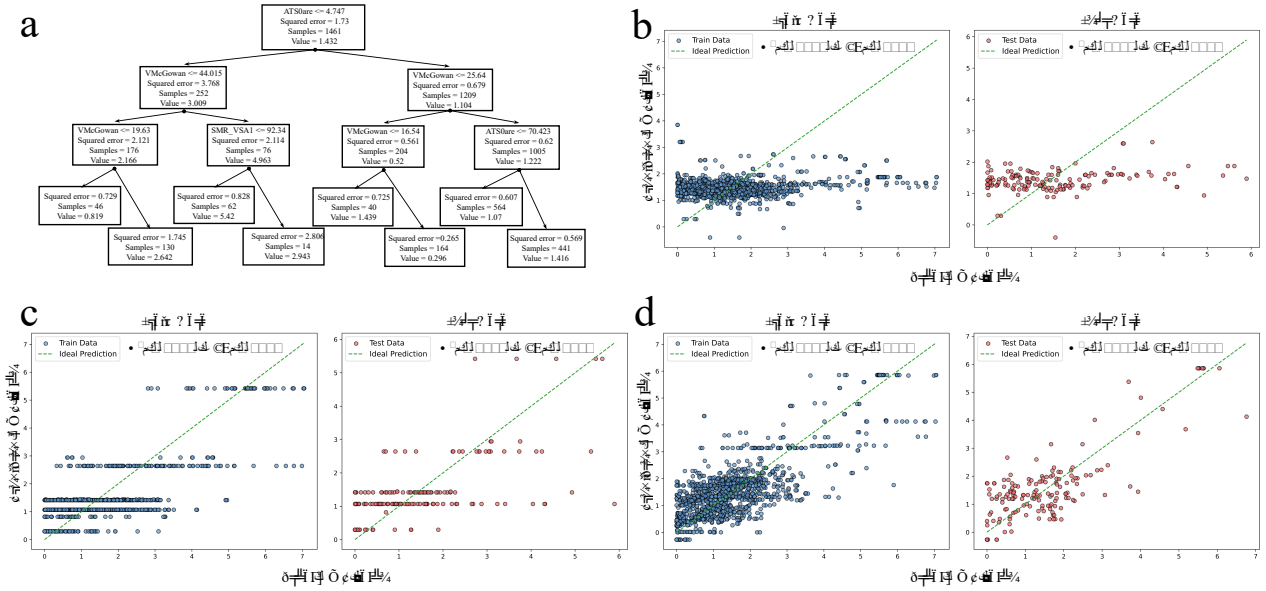

Fig. 3: (a) Visualization of the decision tree model trained on key features for GWP prediction. (b) Prediction results for train and test datasets using the linear regression model. (c) Prediction results for train and test datasets using the decision tree model. (d) Prediction results for train and test datasets using the KAN model.

characteristic equation, exhibited a strong positive correlation with GWP. This suggests that larger molecular structures, often associated with higher molecular weights, tend to have prolonged atmospheric lifetimes and greater radiative forcing potential, leading to increased environmental impact. In addition to these primary descriptors, SMR_VSA1, SMR, and SSE ranked as the third, fourth, and fifth most significant features in the model's decision-making process. SMR_VSA1, a van der Waals surface area descriptor weighted by molar refractivity, displayed an inverse relationship with GWP. SMR, representing total molecular refractivity, and SSE, measuring solute excess polarizability, both exhibited positive correlations with GWP, indicating that molecular dispersion forces and polarizability significantly influence environmental impact predictions. These findings highlight the interplay between molecular size, electronic properties, and intermolecular interactions in shaping GWP. Notably, the high correlation between ATS0are and other key descriptors, such as SMR (0.91) and SSE (0.99), suggests that electronegativity-driven features strongly influence GWP predictions. This reinforces the need for white box approaches like KAN to provide explicit, interpretable equations linking molecular descriptors to GWP, ensuring greater transparency in sustainability assessments.

White-Box Modelling

A white-box modelling approach is proposed to predict GWP scores of chemicals and processes, utilizing the interpretability functions of KAN model.[10, 11] To benchmark the KAN model, we conducted a comparative analysis with two widely used interpretable models: linear regression and decision tree models. To enhance model interpretability, reducing the number of features is as critical as selecting an appropriate model. Therefore, we used the top 20 features identified in **Fig. 2** as inputs for training the white-box models. These features accounted for 87% of the total feature importance, while the remaining features each contributed no more than 1.7%, indicating relatively small influence on model training.

The fundamental benchmark, the linear regression model, produced poor predictive performance, with R^2 scores of 0.0598 and 0.0621, and MSE values of 1.6266 and 1.6185 for the training and test datasets, respectively, as shown in **Fig. 3**(b). The mathematical formulation derived from the linear regression model for a given molecule m is provided in Eq. (1):

$$GWP_m = \sum_{i=1}^{20} a_i \cdot x_{m,i} + b \qquad (1)$$

where $x_{m,i}$ represents chemical descriptors, a_i denote regression coefficients, and b is the intercept term representing the baseline GWP. The coefficients derived from the model are summarized in **Table 2**.

Among the descriptors, SpMax_Dzpe has the highest positive coefficient, suggesting that the maximum atomic polarizability along the principal moment of inertia axis is strongly associated with increased GWP. Sse and VMcGowan also exhibit notable positive influences, indicating that solute excess polarizability and molecular volume play key roles in determining GWP. Conversely, SMR_VSA1 has the most significant negative impact, implying that molecular refractivity-weighted van der Waals surface area contributes to lowering GWP.

These observations highlight the limitations of linear

Table 2: Coefficients of the linear regression model for GWP prediction.

Coefficient	Chemical descriptor	Value
a_1	ATS0are	0.007946
a_2	VMcGowan	0.151153
a_3	SMR_VSA1	-0.800855
a_4	SMR	0.092298
a_5	Sse	0.235614
a_6	BCUTv-1l	-0.000368
a_7	ETA_alpha	-0.637297
a_8	AATSC1c	0.003819
a_9	SpMax_Dzpe	0.745527
a_{10}	SlogP_VSA2	-0.003849
a_{11}	SM1_Dzpe	-0.010904
a_{12}	AMID_h	-0.042039
a_{13}	BCUTse-1l	0.060728
a_{14}	SlogP_VSA10	-0.042921
a_{15}	BCUTi-1h	-0.054201
a_{16}	AXp-0dv	-0.002186
a_{17}	Mare	-0.119402
a_{18}	BCUTpe-1h	-0.060882
a_{19}	AETA_beta	0.047351
a_{20}	SpMAD_Dzse	0.062921
b	-	1.610488

regression in capturing complex, non-linear relationships between molecular descriptors and GWP. Additionally, the presence of collinear features, such as VMcGowan and Sse, suggests potential redundancies in the feature set, which may further restrict the model's predictive capabilities. While the linear regression model offers simplicity and interpretability, its performance remains inadequate for accurate GWP estimation, underscoring the need for more sophisticated modeling approaches that can effectively capture intricate molecular interactions.

The decision tree model, which allows for more complex yet interpretable modelling, yielded the results visualized in **Fig. 3**(a). The boxed structures in the figure represent the decision criteria at each node of the RF model. The tree comprises 15 nodes, with the root node split on ATS6ave, indicating its dominant influence on GWP prediction. Additional splits based on VMcGowan and SMR_VSA1 further indicate the hierarchical importance of these features. As the depth of the tree increases, prediction error decreases, and terminal nodes provide distinct GWP values. However, as depicted in **Fig. 3**(c), the model's predictive performance remains limited, with R^2 scores of 0.5978 and 0.3711, and MSE values of 0.6958 and 1.0852 for the training and test datasets, respectively. While the decision tree model captures non-linear patterns to some extent, its predictive performance still falls short of effectively modelling the complex, non-linear behaviour of chemical properties governing GWP. This demonstrates the inherent trade-off between interpretability and predictive accuracy in traditional white-box models.

To address these limitations, we adopted the KAN framework, widely recognized for its capability to represent symbolic expressions, to develop a GWP prediction model.[19] The KAN model captures non-linear relationships between features and GWP by transforming these relationships into interpretable standard functions such as exponential, polynomial, and trigonometric functions. However, if the fitted model is inherently linear, it can be transformed into a linear model through symbolic transformation, distinguishing KAN from conventional non-linear regression models. The development process of this white-box KAN model, including key steps such as fitting, pruning, fine-tuning, and symbolic simplification. In particular, symbolic simplification is a critical process for ensuring model interpretability. During this process, the model fits various candidate symbolic functions—such as sine, cosine, exponential, logarithmic, and power functions—and systematically selects the form that best represents the underlying non-linear relationship in the data. This automated selection ensures that the chosen functions accurately capture the data's structural patterns while maintaining model interpretability.[10]

As shown in **Fig. 3**(d), the model achieved predictive performance with R^2 scores of 0.6287 and 0.6433, and MSE values of 0.6352 and 0.6716 for the training and test datasets, respectively. Remarkably, despite being a fully interpretable white-box model, its performance is comparable to or even exceeds that of black-box models developed in previous studies, which achieved accuracies of 0.63 and 0.48.[3, 5] The mathematical expression for the resulting white-box model is presented in Eq. (2) as follows:

$$\mathrm{GWP_m} = \varepsilon \cdot \cos\left(\sum_{i=1}^{15} g_i(x_{m,i}) + \eta\right) + \iota \qquad (2)$$

where m denotes the molecule, and $x_{m,i}$ represents the input molecular descriptors specific to molecule m. the parameters ε, η, and ι are set to 5.34842491149902, 8.4815278578314, and 4.47257137298584, respectively. Each function $\mathrm{g_i}(x_{m,j})$ is defined in a general form in Eq. (3) as follows:

$$g_i(x_{m,i}) = \alpha_i \cdot func_i(\beta_i \cdot x_{m,i} + \gamma_i) \qquad (3)$$

where $x_{m,i}$ represents the i-th molecular descriptor of molecule m. The parameter α_i is a scaling coefficient that determines the magnitude of the transformation, while $\beta_i$and γ_i adjust the input values of the function. The function $func_i$ represents a non-linear transformation applied to the descriptor, which can take forms such as sine, cosine, hyperbolic tangent, absolute value, or power functions. The specific functional forms and parameter values used in this model are summarized in **Table 3**. Among the initial 20 descriptors, five—Sse, AATSC1c, SlogP_VSA10, AETA_beta, and SpMAD_Dzse—were excluded during the pruning process of the KAN model. This process systematically removes features that contribute minimally to

the model's predictive performance or exhibit high redundancy with other descriptors.

The function $g_1(x_{m,1})$ is formulated as a scaled hyperbolic tangent function. This structure suggests that $g_1(x_{m,1})$ primarily captures the non-linear effects of ATS0are on GWP, applying a transformation that emphasizes variations within a specific range, while mitigating extreme values through the saturation properties of the hyperbolic tangent function.

The incorporation of a cosine transformation in Eq. (2) allows the model to capture non-linear saturation effects in GWP contributions. This aligns with chemical intuition, reflecting the stabilising effects of certain structural features as they grow larger or more complex. These transformations enable the KAN model to generalise domain-specific knowledge while maintaining interpretability, bridging the gap between predictive accuracy and interpretability. By integrating these features and their transformations, the KAN model delivers a robust framework for understanding the interplay between molecular descriptors and GWP. The result is a white-box model that combines predictive performance with interpretability, adhering to established chemical principles to provide precise forecasts of environmental impact.

Table 3: Functional transformations of molecular descriptors in the KAN model

g_i	$x_{m,i}$	$func_i$	α_i	β_i	γ_i
g_1	ATS0are	tanh	0.090	2.205	3.226
g_2	BCUTv-1l	sin	0.199	2.123	-6.409
g_3	BCUTse-1l	sin	0.185	0.737	5.968
g_4	Mare	sin	-0.192	0.914	-2.037
g_5	BCUTpe-1h	sin	-0.146	7.587	-8.201
g_6	ETA_alpha	cos	0.032	8.995	0.996
g_7	SMR_VSA1	cos	-0.038	2.801	-3.805
g_8	BCUTi-1h	cos	-0.084	2.621	-1.199
g_9	SpMax_Dzpe	cos	-0.051	9.357	-5.015
g_{10}	SlogP_VSA2	tanh	0.186	0.903	-1.762
g_{11}	SMR	Abs	0.007	6.364	-5.575
g_{12}	SM1_Dzpe	Abs	-0.014	9.182	2.453
g_{13}	VMcGowan	Power (3rd)	0.016	-1	0.007
g_{14}	AXp-0dv	Power (4th)	0.003	-1	0.719
g_{15}	AMID_h	Power (2nd)	0.014	1	0.083

CONCLUSIONS

This study highlights the effectiveness of KAN based white-box modeling in GWP prediction. By leveraging the KAN model, we developed a predictive framework that effectively captures complex molecular relationships while maintaining interpretability through symbolic equations. Our findings emphasize the crucial role of physicochemical descriptors, particularly molecular electronegativity and volume, in determining GWP values, reinforcing the feasibility of using purely molecular information for environmental impact assessments.

The developed KAN model offers a structured, explainable approach enhancing transparency in environmental impact modeling. By deriving explicit symbolic equations, the model ensures greater interpretability and supports informed decision-making in the sustainable design and evaluation of chemical processes. This interpretability makes the model highly applicable in industrial and regulatory contexts where transparency is essential. Furthermore, our results demonstrate that KAN achieves predictive accuracy comparable to conventional deep learning models while addressing key interpretability challenges in machine learning-based GWP predictions.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the financial support from the King's College London Net Zero Centre Ph.D. Scholarship scheme.

REFERENCES

1. L. Torrente-Murciano, J.B. Dunn, et al., Nat. Chem. Eng., 1 (2024) 18-27.
2. M. Peplow, Nature, 603 (2022) 780-783.
3. R. Song, A.A. Keller, S. Suh, Environmental science & technology, 51 (2017) 10777-10785.
4. Y. Sun, X. Wang, N. Ren, Y. Liu, S. You, Environmental Science & Technology, 57 (2022) 3434-3444.
5. X. Zhu, C.-H. Ho, X. Wang, ACS Sustainable Chemistry & Engineering, 8 (2020) 11141-11151.
6. S.J. Silva, C.A. Keller, Artificial Intelligence for the Earth Systems, 3 (2024) e230045.
7. K. Letrache, M. Ramdani, in: 2023 14th International Conference on Intelligent Systems: Theories and Applications (SITA), IEEE, 2023, pp. 1-8.
8. G.P. Wellawatte, P. Schwaller, arXiv preprint arXiv:2311.04047, (2023).
9. S. Lundberg, arXiv preprint arXiv:1705.07874, (2017).
10. Z. Liu, P. Ma, Y. Wang, W. Matusik, M. Tegmark, arXiv preprint arXiv:2408.10205, (2024).
11. Z.Liu, Y. Wang, et al., arXiv preprint arXiv:2404.19756, (2024).
12. J.L. Durant, B.A. Leland, et al., J. Chem. Inf. Comput., 42 (2002) 1273-1280.
13. H. Moriwaki, Y.-S. Tian, N. Kawashita, T. Takagi, Journal of cheminformatics, 10 (2018) 1-14.
14. L. Breiman, Machine learning, 45 (2001) 5-32.
15. T. Chen, C. Guestrin, in: Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, 2016, pp. 785-794.
16. D. Svozil, V. Kvasnicka, J. Pospichal, Chemometrics and intelligent laboratory systems, 39 (1997) 43-62.

17. F. Pedregosa, G. Varoquaux, et. al., the Journal of machine Learning research, 12 (2011) 2825-2830.
18. A. Paszke, S. Gross, et. al., Advances in neural information processing systems, 32 (2019).
19. B.C. Koenig, S. Kim, S. Deng, Computer Methods in Applied Mechanics and Engineering, 432 (2024) 117397.