Coordination Chemistry Reviews 534 (2025) 216584

Contents lists available at ScienceDirect

Coordination Chemistry Reviews

journal homepage: www.elsevier.com/locate/ccr

Correspondence

Reevaluating feature importance in gas–solid interaction predictions: A call for robust statistical methods

ARTICLE INFO

Keywords:
Feature importance
Machine learning
Gas–solid interactions
Statistical methods
Model bias
Metal–organic frameworks

ABSTRACT

Liu et al. conducted an insightful investigation into feature importance analysis for predicting CH4 adsorption isotherms in metal–organic frameworks (MOFs), revealing key geometric features that influence model predictions. While their use of advanced machine learning techniques, including neural networks and extra tree regression (ETR), achieved notable accuracy, concerns arise regarding the model-specific biases in feature importance metrics. This paper critically evaluates these metrics, highlighting the risks of misinterpretation due to the lack of ground truth validation. We advocate for the adoption of bias-free statistical methods, such as Spearman's rank correlation and Kendall's tau, which offer a more reliable framework for assessing feature importance. Implementing these approaches could enhance the understanding of gas–solid interactions and improve the reliability of machine learning applications in this domain.

Data analysis machine learning tools can be broadly categorized into two methods: those that rely on ground truth values and those that do not. In supervised machine learning, ground truth values enable the validation of target prediction accuracy, whereas in unsupervised techniques—such as feature importance evaluation, feature selection, feature reduction, and clustering—these values are absent, necessitating extra caution during analysis. Consequently, high target prediction accuracy does not automatically ensure reliable feature importances, as accuracy validation in this context is limited solely to predicting targets and does not extend to verifying the significance of individual features. In contrast, statistical analysis leverages three key elements to accurately determine the true associations between targets and features in the absence of ground truth values: the underlying data distribution, the statistical relationships between variables, and the validation of these relationships through *p*-values. These methods provide a more robust framework for inference compared to feature importances derived from machine learning models, which often do not incorporate these critical statistical components.

Liu et al. conducted an in-depth study on feature importance analysis in gas–solid interaction materials and devices, specifically targeting the prediction of CH4 adsorption isotherms in metal–organic frameworks (MOFs) [1]. They developed neural network protocols that demonstrated remarkable accuracy, achieving an R^2 value of 0.992 for the geometric model. Their analysis of feature importance revealed three pivotal geometric factors—pore fraction, surface area, and density—that significantly influenced the model's predictive capabilities [1]. This rigorous approach not only enhances our understanding of gas adsorption behavior but also underscores the critical role these geometric features play in optimizing MOF design for improved performance [1].

Furthermore, Liu et al. optimized their machine learning approach using tree-integrated extra tree regression (ETR), which achieved a root mean square error (RMSE) as low as 0.24 eV. Their subsequent feature importance analysis revealed that elemental group characteristics from the periodic table, surface energy, and melting point were essential descriptors for predicting gas adsorption behavior [1]. By employing the random forest (RF) algorithm for a deeper analysis, they found that the electric dipole moment proved to be particularly effective in enhancing predictions related to adsorption processes. However, the models utilized, including support vector machines and neural networks, operated as black boxes, complicating the direct interpretation of individual feature contributions [1]. Liu et al. introduced SHAP with these models.

While Liu et al. provided valuable insights into machine learning applications for gas–solid interaction materials and devices, this paper raises significant concerns regarding the reliance on feature importance metrics derived from models such as neural networks, extra tree regression (ETR), and random forests (RF). These techniques are inherently model-specific, and their associated non-negligible biases can lead to misleading conclusions. Importantly, while predictions generated by machine learning can be validated against ground truth values, the metrics for feature importance lack similar validation, increasing the risk of misinterpretation. It is crucial for Liu et al. to recognize that a high level of target prediction accuracy does not necessarily imply the reliability of feature importance metrics.

Different modeling techniques employ distinct methodologies for calculating feature importance, resulting in varied outcomes. This variability highlights the potential biases inherent in these metrics, especially when ground truth values are unavailable. Concerns about bias in feature importance assessments from machine learning models have been extensively documented in over 100 peer-reviewed articles [2–5]. Furthermore, approaches like SHAP (SHapley Additive exPlanations) are particularly susceptible to this issue, as they rely solely on the underlying model. Consequently, SHAP inherits and even amplifies the biases present in the machine learning models themselves, which can lead to misleading conclusions [6–10]. This underscores the necessity for a careful examination of feature importance metrics and the implementation of more robust and unbiased statistical methods in future

https://doi.org/10.1016/j.ccr.2025.216584
Received 22 December 2024; Received in revised form 20 February 2025; Accepted 23 February 2025

analyses.

Although SHAP is designed as a model-agnostic tool to explain predictions by calculating the marginal contribution of each feature, its outputs are inherently tied to the behavior of the underlying machine learning model. In practice, when you invoke a function such as 'explain = SHAP(model)', the resulting SHAP values directly reflect the biases, limitations, and feature importance as determined by the provided model. If the model itself is biased or has generated skewed feature importances—often due to issues like the absence of ground truth for validation—these biases will be inherited (and potentially amplified) by the SHAP explanation [6–10]. In contrast, our proposed statistical methods, illustrated by the function 'explain = STATS(data)', operate directly on the data rather than on a model's outputs. This approach addresses feature relationships using fundamental statistical principles (such as data distribution, variable relationships, and *p*-values), thereby offering a framework that is not susceptible to the biases of any specific model.

In light of these issues, this paper advocates for the use of bias-free statistical methods, such as Spearman's rank correlation with *p*-values [11,12] and Kendall's tau with p-values [13,14], both of which are nonlinear and nonparametric approaches. These methods offer a more robust and reliable framework for evaluating feature importance while minimizing biases associated with traditional machine learning evaluations. By adopting such approaches, researchers can achieve clearer and more accurate insights into the factors influencing gas–solid interactions.

While Spearman's rank correlation and Kendall's tau, along with their accompanying *p*-values, effectively measure nonlinear and nonparametric associations between two variables, they are not designed to capture the nuanced contributions of feature variables to model predictions. In practice, the correlation between a feature and the prediction outcome may not align with the feature's actual influence when integrated into a predictive model, largely due to interactions and complex dependencies within the model itself. The ultimate goal of assessing feature importance is to accurately quantify the true associations between variables. However, because models often lack ground truth values and may incorporate biases, the feature importances derived from them can be inherently skewed. Our proposed approach utilizes robust statistical methods to calculate true associations, offering an evaluation that is not tainted by the biases present in model outputs. While this paper acknowledges the value of model-based target predictions, it argues that true associations should be assessed through these unbiased statistical measures rather than relying solely on potentially misleading feature importances from the model.

To clarify, Liu et al. provide a comprehensive review of existing machine learning models used for predicting gas adsorption rather than developing these models themselves. Their work synthesizes and evaluates various approaches from the literature, highlighting current trends, challenges, and opportunities in the field. We have updated the manuscript accordingly to accurately reflect that Liu et al.'s contribution lies in their critical review of the state-of-the-art rather than in model development.

While acknowledging that machine learning is effectively used for target prediction in gas catalysis, this paper primarily raises critical concerns about the reliance on interpretative methods such as feature importance and SHAP, highlighting their potential biases and limitations.

Funding

This research has no fund.

Ethics approval

Not applicable.

Consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and material

Not applicable.

Code availability

Not applicable.

Authors' contribution

Yoshiyasu Takefuji completed this research and wrote this article.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

References

[1] P. Liu, X. Zhu, X. Ran, H. Bi, X. Huang, N. Gu, Machine learning for gas–solid interaction materials and devices, Coord. Chem. Rev. 524 (2025) 216329, https://doi.org/10.1016/j.ccr.2024.216329.

[2] A. Fisher, C. Rudin, F. Dominici, All models are wrong, but many are useful: learning a variable's importance by studying an entire class of prediction models simultaneously, J. Mach. Learn. Res. 20 (2019) 177.

[3] Demircioğlu A. Measuring the bias of incorrect application of feature selection when using cross-validation in radiomics. Insights Imaging. 2021;12(1):172. Published 2021 Nov 24. doi:https://doi.org/10.1186/s13244-021-01115-1.

[4] J.L. Cross, M.A. Choma, J.A. Onofrey, Bias in medical AI: implications for clinical decision-making, PLoS Digit Health 3 (11) (2024) e0000651. Published 2024 Nov 7, https://doi.org/10.1371/journal.pdig.0000651.

[5] T.T. Nguyen, J.Z. Huang, T.T. Nguyen, Unbiased feature selection in learning random forests for high-dimensional data, Sci. World J. 2015 (2015) 471371, https://doi.org/10.1155/2015/471371.
[6] B. Bilodeau, N. Jaques, P.W. Koh, B. Kim, Impossibility theorems for feature attribution, Proc. Natl. Acad. Sci. U.S.A. 121 (2) (2024) e2304406120, https://doi.org/10.1073/pnas.2304406120.
[7] X. Huang, J. Marques-Silva, On the failings of Shapley values for explainability, Int. J. Approx. Reason. 171 (2024) 109112, https://doi.org/10.1016/j.ijar.2023.109112.
[8] M.A. Lones, Avoiding common machine learning pitfalls, Patterns 5 (10) (2024) 101046, https://doi.org/10.1016/j.patter.2024.101046.
[9] C. Molnar, et al., General pitfalls of model-agnostic interpretation methods for machine learning models, in: A. Holzinger, R. Goebel, R. Fong, T. Moon, K. R. Müller, W. Samek (Eds.), xxAI - Beyond Explainable AI. xxAI 2020, Lecture Notes in Computer Science, vol. 13200, Springer, 2022, https://doi.org/10.1007/978-3-031-04083-2_4.
[10] I. Kumar, C. Scheidegger, S. Venkatasubramanian, S. Friedler, Shapley residuals: quantifying the limits of the shapley value for explanations, Adv. Neural Inf. Process. Syst. 34 (2021) 26598–26608.
[11] H. Yu, A.D. Hutson, A robust Spearman correlation coefficient permutation test, Commun. Stat. Theory Methods 53 (6) (2024) 2141–2153, https://doi.org/10.1080/03610926.2022.2121144.
[12] S.K. Eden, C. Li, B.E. Shepherd, Nonparametric estimation of Spearman's rank correlation with bivariate survival data, Biometrics 78 (2) (2022) 421–434, https://doi.org/10.1111/biom.13453.
[13] J.H. Wang, Y.H. Chen, Network-adjusted Kendall's tau measure for feature screening with application to high-dimensional survival genomic data, Bioinformatics 37 (15) (2021) 2150–2156, https://doi.org/10.1093/bioinformatics/btab064.
[14] N. Ouachene, T.S. Kiessé, M.S. Corson, Using conditional Kendall's tau estimation to assess interactions among variables in dairy-cattle systems, Agric. Syst. 220 (2024) 104089, https://doi.org/10.1016/j.agsy.2024.104089.

Yoshiyasu Takefuji[1]
Faculty of Data Science, Musashino University, 3-3-3 Ariake Koto-ku, Tokyo 135-8181, Japan
E-mail address: takefuji@keio.jp.

[1] **ORCID:** 0000-0002-1826-742X.