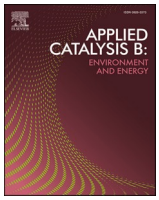
Applied Catalysis B: Environment and Energy 368 (2025) 125145

Contents lists available at ScienceDirect

Applied Catalysis B: Environment and Energy

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



Reevaluating feature importance in machine learning models for CO2 photoreduction: A statistical perspective

ARTICLE INFO

Keywords:
CO2 photoreduction
Feature importance
Machine learning
Statistical analysis
Model biases
VIF analysis

ABSTRACT

Chen et al. have advanced the theoretical design of dual-site metallo-covalent organic frameworks for enhancing CO2 photoreduction into C2H4 using various machine learning algorithms. While they demonstrated high predictive accuracy using a stacking approach with seven selected algorithms, this study emphasizes the potential biases in feature importance derived from these models. It argues for the necessity of computing unbiased feature importances and highlights the complications posed by different methodologies across models. Further, it recommends robust statistical techniques, such as Spearman's correlation and Kendall's tau, to improve interpretability and validity. Addressing collinearity through Variance Inflation Factor (VIF) analysis is also crucial. These steps aim to deepen understanding and optimize machine learning applications for carbon capture and utilization.

Chen et al. developed machine learning models for the theoretical design of dual-site metallo-covalent organic frameworks, with the goal of enhancing the efficiency of CO2 photoreduction into C2H4 [1]. In their study, they employed a diverse array of algorithms commonly utilized for regression problems. These algorithms included decision tree regression (DTR), elastic net regression (ENR), extra trees regression (ETR), K-nearest neighbor regression (KNNR), least absolute shrinkage and selection operator (LASSO), linear regression (LR), support vector regression (SVR), gradient boosting regression (GBR), XGBoost regression (XGBR), and random forest regression (RFR) [1].

To effectively leverage the characteristics of stacking algorithms, the researchers conducted comprehensive experiments aimed at identifying suitable individual and meta-learners [1]. The results revealed the R^2 scores and root mean square error (RMSE) values derived from the stacking approach, showcasing its predictive capabilities. Ultimately, seven machine learning algorithms that exhibited strong performance were selected as base learners to train and predict outcomes on the dataset, with the training and testing sets divided in a 75:25 ratio. A thorough comparison of the algorithms' running times, R^2 scores, and RMSE values led to the selection of ETR as the meta-learner, while DTR, ENR, ETR, KNNR, GBR, XGBR, and RFR formed the base learners to build the stacking model. Furthermore, a feature importance analysis was conducted based on the stacking model's results, providing insights into the predictors influencing CO2 photoreduction efficiency [1].

However, while this paper highlights the innovative high accuracy of the theoretical models, it also raises significant concerns regarding the reliance on feature importances derived from machine learning models. Distorted feature importances can lead to erroneous conclusions that undermine the theoretical design intentions. To enhance the efficiency of CO2 photoreduction, it is essential for Chen et al. to compute bias-free or bias-less feature importances. Unlike target predictions, which have ground truth values to validate accuracy, feature importances often lack comparable benchmarks for validation. Consequently, different machine learning models utilize varied methodologies for calculating feature importances, a point illustrated by Chen et al.'s findings that distinct models generate disparate feature importances, indicative of potential non-negligible biases.

The challenge of accurately assessing feature importance is further complicated by findings from over 100 peer-reviewed articles, which highlight non-negligible biases in feature importances computed from machine learning models [2–5]. While R^2 scores and root mean square error (RMSE) values provide valuable insights into how well data fit the chosen models, they do not offer a comprehensive understanding of target accuracy; specifically, they do not convey critical information about sensitivity, specificity, or overall accuracy [6].

It is essential to recognize that achieving high accuracy in target predictions does not necessarily guarantee the reliability of feature importance evaluations. In the absence of established ground truth values for feature importance, researchers must carefully evaluate three key aspects: the underlying data distribution, the statistical relationships among variables, and the validation of these relationships through p-values.

By addressing these factors, researchers can improve the interpretability and robustness of their analyses, ultimately leading to more accurate assessments of feature importance. Recognizing the limitations of common metrics, as well as the nuances of the underlying data, is vital for advancing the reliability of conclusions drawn from machine learning models.

To achieve a rigorous theoretical analysis that accurately elucidates the genuine associations between the target variable and the features, it is essential to employ robust statistical methodologies that minimize potential distortions. This paper advocates the use of nonlinear and nonparametric techniques [7–10], such as Spearman's correlation with p-values or Kendall's tau with p-values. These methods provide a more nuanced understanding of relationships among variables without the assumptions inherent in linear models.

https://doi.org/10.1016/j.apcatb.2025.125145
Received 28 December 2024; Received in revised form 22 January 2025; Accepted 8 February 2025
Available online 12 February 2025

Prior to implementing these recommended approaches, it is imperative to address issues of collinearity and feature interactions, as such factors can significantly inflate correlation values and distort results. Conducting a Variance Inflation Factor (VIF) analysis is vital for identifying and alleviating the effects of these troublesome features [11]. VIF is a valuable tool for detecting collinearity and interactions among features. It is advisable to apply VIF prior to employing statistical methods in order to remove features exhibiting collinearity and interactions. This proactive step helps mitigate feature inflation and enhances the robustness of the analysis. This approach not only enhances the reliability of feature importance assessments but also enriches the overall interpretability of results. By adopting these best practices, researchers can significantly advance the understanding and optimization of CO2 photoreduction processes using machine learning techniques, ultimately contributing to more effective strategies for carbon capture and utilization.

Issues related to feature importance are prevalent across various fields, including medicine, multi-omics, pharmacology, chemistry, material science, and beyond. Many researchers are often unaware that ground truth values are absent in feature importance calculations from machine learning models, leading to potential confusion with prediction accuracy. It is crucial to understand that high prediction accuracy does not necessarily ensure the reliability of feature importance metrics.

Ethics approval

Not applicable.

Authors' contributions

Yoshiyasu Takefuji completed this research and wrote this article.

Consent to participate

Not applicable.

Consent for publication

Not applicable.

Code availability

Not applicable.

Funding

This research has no fund.

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] K.-X. Chen, Y. Wang, Y.-Q. Dong, Z.-H. Zhao, Y. Zang, G.-L. Zhang, Y. Liu, F.-M. Zhang, Theoretical design of dual-site metallo-covalent organic frameworks for efficient CO2 photoreduction into C2H4, Appl. Catal. B Environ. 365 (2025) 124933, https://doi.org/10.1016/j.apcatb.2024.124933.

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

[4] K.L. Devick, I. Zaniletti, D.R. Larson, D.G. Lewallen, D.J. Berry, H. Maradit Kremers, Avoiding systematic bias in orthopedics research through informed variable selection: a discussion of confounders, mediators, and colliders, J. Arthroplast. 37 (10) (2022) 1951–1955, https://doi.org/10.1016/j.arth.2022.05.027.

[5] Y.H. Yu, K.B. Filion, L.M. Bodnar, et al., Visualization tool of variable selection in bias-variance tradeoff for inverse probability weights, Ann. Epidemiol. 41 (2020) 56–59, https://doi.org/10.1016/j.annepidem.2019.12.006.

[6] D.L. Alexander, A. Tropsha, D.A. Winkler, Beware of R^2: simple, unambiguous assessment of the prediction accuracy of QSAR and QSPR models, J. Chem. Inf. Model 55 (7) (2015) 1316–1322, https://doi.org/10.1021/acs.jcim.5b00206.

[7] 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.

[8] 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.

[9] S. Chen, A. Ghadami, B.I. Epureanu, Practical guide to using Kendall's τ in the context of forecasting critical transitions, Published 2022 Jul 27. R. Soc. Open Sci. 9 (7) (2022) 211346 https://doi.org/10.1098/rsos.211346.

[10] X. Qi, Y. Lu, Y. Shi, H. Qi, L. Ren, A deep neural network prediction method for diabetes based on Kendall's correlation coefficient and attention mechanism, Published 2024 Jul 2. PLoS One 19 (7) (2024) e0306090 https://doi.org/10.1371/journal.pone.0306090.

[11] J. Jacob, R. Varadharajan, Robust variance inflation factor: a promising approach for collinearity diagnostics in the presence of outliers, Sankhya B 86 (2024) 845–871, https://doi.org/10.1007/s13571-024-00342-y.

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