*Interdisciplinary Materials*

WILEY

Interdisciplinary Materials Open Access

COMMENT OPEN ACCESS

# Reassessing Machine Learning Techniques for Electrocatalyst Design: A Call for Robust Methodologies

Yoshiyasu Takefuji

Faculty of Data Science, Musashino University, Tokyo, Japan

**Correspondence:** Yoshiyasu Takefuji (takefuji@keio.jp)

**Received:** 10 July 2025 | **Revised:** 10 July 2025 | **Accepted:** 13 August 2025

**Funding:** The author received no specific funding for this work.

**Keywords:** electrocatalysts | feature selection | machine learning | SHAP | Spearman's correlation

Gu et al. conducted a comprehensive survey on the design and application of electrocatalysts powered by machine learning techniques [1]. They presented a novel approach that utilizes Artificial Neural Networks (ANN) in conjunction with the SHAP (SHapley Additive exPlanations) method to optimize membrane electrode assemblies. The ANN model demonstrated high accuracy in predicting key performance metrics, achieving root mean square error (RMSE) values of 43.536 $mW\,cm^{-2}$ for power density and 0.070 $gPt\ kW^{-1}$ for platinum utilization. Additionally, the SHAP method was employed to identify the most influential features affecting the target outputs, providing valuable insights into the optimization process [1].

However, this paper raises significant theoretical and empirical concerns regarding the use of ANN in conjunction with SHAP due to the model-specific nature of these techniques, which can lead to erroneous interpretations. It appears that Gu et al. may not fully grasp the fundamental principles underlying machine learning. In supervised machine learning models like ANN, two types of accuracy are crucial: target prediction accuracy and feature importance reliability. While target prediction accuracy can be validated against known ground truth values, the derived feature importances from models lack equivalent ground truth for validation. As a result, achieving high target prediction accuracy does not ensure that the feature importances are also reliable, since there are no established ground truth values for these features. The function call "explain = SHAP(model)" further indicates that SHAP may inherit and potentially amplify any biases present in the feature importances derived from the underlying model (ANN), leading to misleading interpretations of the results [2–5]. This highlights the importance of critically evaluating both the predictions and the interpretability provided by model-agnostic methods like SHAP.

In light of these concerns, the paper advocates for a more robust and multifaceted approach utilizing unsupervised machine learning techniques, such as Feature Agglomeration (FA) and Highly Variable Gene Selection (HVGS). FA is a dimensionality reduction technique that aggregates similar features, thereby simplifying the data set and reducing noise, which can enhance the interpretability of the model and the reliability of its predictions. HVGS focuses on selecting a subset of features that exhibit significant variability across samples, ensuring that only the most informative features are retained for further analysis.

Following the feature selection process, the authors suggest employing nonlinear nonparametric statistical methods, such as Spearman's correlation, to assess the relationships between features and outcomes. Spearman's correlation evaluates the strength and direction of the association between ranked variables, making it particularly useful in identifying monotonic relationships that do not necessarily follow a linear pattern. The accompanying $p$-values provide a measure of statistical significance, offering insights into the reliability of these correlations. By leveraging these advanced methods, researchers can achieve a deeper understanding of the key factors influencing electrocatalyst performance, while also mitigating the risks associated with relying solely on model-specific interpretations.

## Author Contributions

Yoshiyasu Takefuji completed this study and wrote this article.

## Ethics Statement

The author has nothing to report.

## Consent

The author has nothing to report.

## Conflicts of Interest

The author declares no conflicts of interest.

## Data Availability Statement

The author has nothing to report.

## References


1. Y. Gu, H. Zhang, Z. Xu, et al., "Design and Application of Electrocatalyst Based on Machine Learning," *Interdisciplinary Materials* 4 (2025): 456–479, https://doi.org/10.1002/idm2.12249.

2. L. Wu, "A Review of the Transition From Shapley Values and SHAP Values to RGE," *Statistics* 1 (2025): 1–23, https://doi.org/10.1080/02331888.2025.2487853.

3. B. Bilodeau, N. Jaques, P. W. Koh, and B. Kim, "Impossibility Theorems for Feature Attribution," *Proceedings of the National Academy of Sciences of the United States of America* 121, no. 2 (2024): e2304406120, https://doi.org/10.1073/pnas.2304406120.

4. X. Huang and J. Marques-Silva, "On the Failings of Shapley Values for Explainability," *International Journal of Approximate Reasoning* 171 (2024): 109112, https://doi.org/10.1016/j.ijar.2023.109112.

5. D. Hooshyar and Y. Yang, "Problems With SHAP and LIME in Interpretable AI for Education: A Comparative Study of Post-Hoc Explanations and Neural-Symbolic Rule Extraction," *IEEE Access* 12 (2024): 137472–137490, https://doi.org/10.1109/ACCESS.2024.3463948.