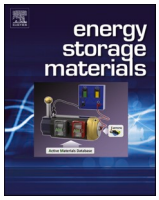
Energy Storage Materials 78 (2025) 104242

Contents lists available at ScienceDirect

Energy Storage Materials

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



Enhancing feature importance analysis in battery research: a statistical methods perspective on machine learning limitations

ARTICLE INFO

Keywords:
Feature importance analysis
Statistical validation
Battery modeling
Machine learning limitations
Nonparametric methods

ABSTRACT

This paper addresses critical concerns related to feature importance analysis in battery research, specifically examining the limitations of machine learning-derived feature importances as reported by Yuan et al. While recent studies have achieved impressive prediction accuracy in battery modeling, this paper underscores that such accuracy does not necessarily ensure the trustworthy interpretation of feature importances. This paper advocates for the adoption of robust statistical methods as a superior alternative to model-derived feature importances, emphasizing three key advantages: the provision of directional information (ranging from -1 to +1), standardized comparison scales, and statistical validation through p-values. To enhance the reliability and interpretability of feature importance analysis, this paper introduces a comprehensive framework that incorporates five nonlinear, nonparametric statistical methods. This approach is designed to enhance the rigor and clarity of feature importance assessments in battery research and related fields.

Accurate analysis is essential for researchers to ensure error-free calculations. The choice of machine learning tools can lead to significantly different outcomes, underscoring the importance of understanding their limitations. Researchers must ascertain whether their calculations have ground truth values for validation. In the absence of ground truth, additional caution is necessary to achieve reliable results. This paper presents a detailed example illustrating the use of ground truth values in the context of feature importance and feature selection, highlighting the potential pitfalls of ignoring this crucial aspect in analytical processes. By emphasizing the importance of validation in feature analysis, this work seeks to promote more robust and trustworthy research practices.

Yuan et al. reviewed and investigated computational understanding and multiscale simulation of secondary batteries [1]. They employed linear, lasso, ridge, and elasticnet regression models to predict nanocage vcell, achieving an r^2 score of 0.99 and a root mean square error (rmse) below 0.05. Among these models, lasso regression achieved the highest prediction accuracy and effective feature selection, attributed to its L1 regularization technique [1].

While Yuan et al. provide a pioneering review of computational models for secondary batteries, this paper highlights significant concerns regarding their approach to feature selection and feature importance as derived from lasso regression. The model-specific nature of these analyses can lead to misleading conclusions, as different models can yield varying feature importances. It is crucial for Yuan et al. to grasp the fundamental theoretical principles of machine learning, particularly the distinction between target prediction accuracy and the reliability of feature importance.

In supervised machine learning, while ground truth values are available for validating prediction accuracy, the same cannot be said for feature importances. As a result, high target prediction accuracy does not necessarily imply that the derived feature importances are trustworthy. In essence, the accuracy of target predictions and the validity of feature importance are separate issues. The lack of ground truth values in feature importance computations introduces inherent biases, potentially leading to flawed conclusions. Numerous peer-reviewed studies—over 100—have documented the presence of significant biases in feature importances generated by machine learning models, underscoring the need for caution in interpreting these results [2–7].

Feature importance, feature selection, and feature reduction techniques lack ground truth values for validation, necessitating extra caution in their analysis. In statistics, when ground truth values are absent, three critical components must be considered: data distribution patterns, statistical relationships between variables, and validation through p-values for statistical significance.

This paper advocates for the adoption of nonlinear and nonparametric robust statistical methods, including spearman's correlation [8], kendall's tau [9], goodman-kruskal gamma [10], somers' d [11], and hoeffding's d [12]. These five statistical methods offer distinct advantages over model-derived feature importances due to their comprehensive analytical capabilities.

A key distinction lies in how these methods and model-derived feature importances represent relationships. Model-derived feature importances typically range from 0 to 1, where values closer to 1 indicate stronger influence or importance of a feature, but without indicating the direction of the relationship. In contrast, statistical methods provide three crucial pieces of information:

First, they provide directional information, offering crucial insights about the nature of relationships between variables. When a positive value is obtained, it indicates that variables move in the same direction, while negative values reveal inverse relationships between variables. This directional information is vital for understanding the underlying relationships in the data.

Second, these methods operate within a standardized range from −1

https://doi.org/10.1016/j.ensm.2025.104242
Received 2 February 2025; Received in revised form 15 March 2025; Accepted 4 April 2025
Available online 5 April 2025

to +1, where −1 indicates a perfect negative correlation, 0 indicates no correlation, and +1 indicates a perfect positive correlation. This standardized scale enables direct comparison across different variables and methods, making the results more interpretable and comparable.

Third, these methods are accompanied by their respective p-values, which provide statistical validation of the results. P-values indicate the probability that the observed relationship occurred by chance, offering a quantitative measure of statistical significance. This statistical validation is crucial for ensuring the reliability of findings and helps researchers distinguish meaningful relationships from random fluctuations in the data.

The combination of directional information, standardized range, and statistical validation through p-values makes these methods particularly valuable for robust data analysis, offering advantages that model-derived feature importances typically cannot provide.

To further advance this work, future studies should explore hybrid modeling approaches that combine advanced machine learning techniques with domain-specific knowledge to achieve both high generalization and high precision under the constraints of limited battery parameters. Furthermore, integrating robust nonlinear and nonparametric strategies is critical not only for capturing the complex dynamics of battery degradation but also for accurately predicting the **state of health (soh)** of lithium batteries across diverse chemical systems. Such approaches would facilitate the development of early warning systems capable of detecting potential hazards, thereby enhancing the safety and reliability of battery operation in real-world applications.

In addition, it is important to note that the feature extraction processes for laboratory data and field data can differ significantly due to their distinct aging mechanisms. Laboratory data, collected under controlled experimental conditions, often display predictable and homogeneous degradation patterns, allowing for more straightforward feature extraction and analysis. In contrast, field data are subject to a wide range of operational stresses, environmental influences, and unmonitored variables, which result in more complex and heterogeneous aging behaviors. Therefore, adapting or developing feature extraction methods that specifically address the noise and variability inherent in field data is essential to ensure accurate modeling and reliable prediction of battery performance in real-world scenarios.

Funding

This research has no fund.

Ethics approval

Not applicable

Consent to participate

Not applicable

Consent for publication

Not applicable

CRediT authorship contribution statement

Yoshiyasu Takefuji: Writing – review & editing, Writing – original draft, Validation, Investigation, Conceptualization.

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.

Availability of data and material

Not applicable

Code availability

Not applicable

References

[1] Y. Yuan, B. Wang, J.-H. Zhang, B. Zheng, S.S. Fedotov, H. Lu, L. Kong, Computational understanding and multiscale simulation of secondary batteries, Energy Storage Materials 75 (2025) 104009, https://doi.org/10.1016/j.ensm.2025.104009.

[2] K. Fang, S. Ma, Analyzing large datasets with bootstrap penalization, Biom J 59 (2) (2017) 358–376, https://doi.org/10.1002/bimj.201600052.

[3] H. Sun, X. Wang, High-dimensional feature selection in competing risks modeling: a stable approach using a split-and-merge ensemble algorithm, Biom J 65 (2) (2023) e2100164, https://doi.org/10.1002/bimj.202100164.

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

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

[6] C. Strobl, A.L. Boulesteix, A. Zeileis, T. Hothorn, Bias in random forest variable importance measures: illustrations, sources and a solution, BMC Bioinformatics 8 (2007) 25, https://doi.org/10.1186/1471-2105-8-25.

[7] T. Salles, L. Rocha, M. Gonçalves, A bias-variance analysis of state-of-the-art random forest text classifiers, Adv Data Anal Classif 15 (2021) 379–405, https://doi.org/10.1007/s11634-020-00409-4.

[8] 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.

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

[10] J. Metsämuuronen, Directional nature of Goodman–Kruskal gamma and some consequences: identity of Goodman–Kruskal gamma and Somers delta, and their connection to Jonckheere–Terpstra test statistic, Behaviormetrika 48 (2021) 283–307, https://doi.org/10.1007/s41237-021-00138-8.

[11] R. Newson, Confidence intervals for rank statistics: somers' D and extensions, Stata J 6 (2006) 309–334.

[12] A. Fujita, J.R. Sato, M.A. Demasi, M.C. Sogayar, C.E. Ferreira, S. Miyano, Comparing Pearson, Spearman and Hoeffding's D measure for gene expression association analysis, J Bioinform Comput Biol 7 (4) (2009) 663–684, https://doi.org/10.1142/s0219720009004230.

Yoshiyasu Takefuji*

Faculty of Data Science, Musashino University, 3-3-3 Ariake Koto-ku, Tokyo 135-8181, Japan

* Corresponding Author at: Yoshiyasu Takefuji, Professor, Faculty of Data Science, Musashino University, 3-3-3 Ariake Koto-ku, Tokyo 135-8181, Japan.

E-mail address: takefuji@keio.jp.