Next Materials 9 (2025) 101182

Contents lists available at ScienceDirect

Next Materials

journal homepage: www.sciencedirect.com/journal/next-materials

Discussion

Enhancing reliability in materials science analyses through understanding machine learning principles

Yoshiyasu Takefuji [*,1,2]

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

ARTICLE INFO

Keywords
Machine learning
Materials science
Linear and nonlinear models
Parametric and nonparametric methods
True associations
Robust statistical techniques

ABSTRACT

In the rapidly evolving field of materials science, a solid understanding of machine learning principles is essential for conducting accurate analyses and deriving meaningful conclusions. Ground truth values are critical for validating accuracy, particularly in supervised machine learning, where they serve as benchmarks for target prediction. However, feature importances derived from models often lack corresponding ground truth values, leading to inherent biases in their interpretation. This paper examines three case studies that illustrate the pitfalls associated with a limited understanding of key machine learning concepts, including linear versus nonlinear models and parametric versus nonparametric approaches. By addressing these foundational elements, researchers can enhance the reliability of their findings when investigating complex relationships within materials data. The issues highlighted emphasize the need for employing error-free and bias-free methodologies, such as robust statistical techniques, to ensure credible outcomes. Ultimately, this work advocates for improved practices in machine learning, which are vital for driving significant advancements in materials science.

1. Introduction

Many researchers are unaware of the widespread misapplication of machine learning tools that occurs when fundamental theoretical and empirical principles are not properly understood. In the era of large language models (LLMs), their application in materials science has become increasingly prevalent. However, current generative AI systems based on LLMs typically achieve approximately 85 % accuracy, leaving a 15 % error rate that necessitates careful scrutiny. This underscores the critical need for users to independently validate generated outcomes using appropriate verification methods. When researchers lack solid grounding in machine learning principles, the reliability of their results may be significantly compromised.

The misapplication of machine learning techniques can be categorized into three distinct types: violation of fundamental assumptions underlying AI tools, challenges related to ground truth in model interpretation, and other critical methodological errors. A crucial distinction exists in validation approaches: while supervised machine learning models utilize ground truth values to validate prediction accuracy, the feature importance rankings derived from these models lack corresponding ground truth mechanisms for accuracy validation. Target prediction accuracy and feature importance reliability are distinct issues. In other words, high prediction accuracy does not guarantee reliable feature importances. The reliability of outcomes is fundamentally determined by the integrity, representativeness, and quality of the datasets employed. To ensure robust results and reliable conclusions, multi-faceted validation approaches incorporating domain knowledge, statistical verification, and cross-model comparison are essential, particularly when interpreting model-derived feature importance in materials science applications.

This paper provides empirical benchmarking of common misapplications in machine learning for materials science. Linear models such as LASSO, despite their widespread use and interpretability advantages, can produce distorted and skewed outcomes due to their inherently linear and parametric nature, potentially leading to erroneous interpretations and conclusions. There are no significant differences between materials science and other fields in this regard, as these methodological limitations apply across scientific domains. To

* Corresponding author.
E-mail address: takefuji@keio.jp.
[1] **ORCID:** 0000-0002-1826-742X.
[2] **According to ScholarGPS,** Yoshiyasu Takefuji holds notable global rankings in several fields. He ranks 54th out of 395,884 scholars in neural networks (AI), 23rd out of 47,799 in parallel computing, and 14th out of 7,222 in parallel algorithms. Furthermore, he ranks highest in AI tools and human-induced error analysis, underscoring his significant contributions to these domains.

https://doi.org/10.1016/j.nxmate.2025.101182
Received 25 January 2025; Received in revised form 20 August 2025; Accepted 3 September 2025

demonstrate this phenomenon, this paper evaluates feature selection quality using the MNIST dataset, comprising 70,000 samples with 784 features (corresponding to 28 × 28 pixel images). Three feature selection methodologies to select top 30 features—LASSO, Mutual Information (MI), and Transfer Entropy (TE)—are systematically compared by mnisttest.py for reproducibility purposes [10]. Performance is evaluated using cross-validation accuracy, with the premise that superior feature selection methods yield higher accuracy scores. MI and TE selections are cross-validated using random forest models trained on the selected features. The results conclusively demonstrate that LASSO significantly underperforms in feature selection compared to MI and TE approaches due to limitations of LASSO linearity. Table 1 presents the comprehensive mean cross-validation accuracy metrics that quantify these performance differences.

Symbolic regression (SR) transforms "black box" machine learning into interpretable expressions for materials science by exploring mathematical relationships in data. Their review examines SR theories, workflows, techniques, codes, and applications, while addressing key challenges: GPU acceleration, transfer learning, balancing accuracy with complexity, integration with large language models for physical interpretability, and multimodal approaches—all crucial for unlocking SR's full potential in materials design and research. Liu et al. applied machine learning methods to materials science and developed frameworks that embed domain knowledge into the modeling process. Their work emphasizes the critical importance of incorporating materials domain expertise throughout machine learning model development.

In the rapidly evolving field of materials science, grasping the fundamental principles of machine learning is essential for advancing research and development. Key concepts such as accuracy, reliability, and predictive power directly impact the analyses performed by researchers. As materials become more complex, the demand for sophisticated data analysis techniques capable of effectively interpreting material properties and behaviors increases. Linear and nonlinear models play a vital role in this process, each offering unique advantages and limitations that can significantly influence research outcomes.

Linear models operate under the assumption of a direct, proportional relationship between input features and output targets, making them relatively straightforward to interpret and apply—for instance, in predicting tensile strength from compositional data. This clarity is especially beneficial when researchers aim to gain preliminary insights into material behavior. Conversely, nonlinear models excel at capturing intricate relationships that linear models may overlook. By leveraging techniques like polynomial regression and neural networks, nonlinear models are particularly valuable in complex scenarios, such as phase transitions and stress-strain relationships, where variable interactions can be multifaceted.

The choice between linear and nonlinear modeling approaches has profound implications for research results [4–6]. While linear models are generally more straightforward and less susceptible to overfitting, nonlinear models can provide deeper insights but often act as "black boxes," raising interpretability concerns. Therefore, a comprehensive understanding of these methodologies is essential for researchers who wish to leverage machine learning effectively in materials science.

In addition to selecting modeling approaches, researchers must differentiate between parametric and nonparametric methods. Parametric methods rely on predefined distributions, allowing for precise estimates when underlying assumptions hold true, while nonparametric approaches offer increased flexibility, accommodating a wider variety of data types and distributions. This choice can significantly affect the validity of research conclusions, underscoring the importance of a nuanced understanding of each method's applicability [7–9].

Table 1
Quality of feature selection with 5-fold cross-validation accuracy.

Method	5-fold cross-validation accuracy
Mutual Information (with RF)	0.8445 ± 0.0023
Transfer Entropy (with RF)	0.8237 ± 0.0021
LASSO (with LASSO classifier)	0.7299 ± 0.0024

This paper presents three illustrative case studies [1–3] that highlight potential pitfalls arising from a lack of comprehension in these critical areas. It particularly emphasizes the implications of distinguishing between scenarios with and without ground truth values for accuracy validation. High prediction accuracy associated with ground truth does not automatically guarantee reliable feature importances, feature selections, or feature reductions. Over 100 peer-reviewed articles highlighted non-negligible biases in feature importances from machine learning models.

As machine learning techniques gain prevalence in the field, researchers frequently face challenges stemming from insufficient familiarity with these foundational concepts. Consequently, this paper advocates for adopting robust, unbiased methodologies that ensure research outcomes reflect true relationships within materials data. By promoting best practices in machine learning, we aim to enhance the reliability and effectiveness of analyses, ultimately paving the way for groundbreaking innovations in material development.

Understanding true associations—genuine relationships between target variables and their features—is crucial for effective predictive modeling. Such associations inform how changes in features can affect outcomes, often pointing to causal connections. For instance, in materials science, the composition of an alloy can profoundly impact its tensile strength, illustrating the important distinction between correlation and causation. However, identifying these relationships is often complicated by noise in datasets that can obscure real correlations. Effective data preprocessing, combined with robust statistical methods, is essential for isolating true associations. Furthermore, genuine relationships should demonstrate consistency across studies, enhancing their credibility and suggesting underlying physical mechanisms.

In conclusion, true associations are characterized by consistency, significance, and reproducibility—critical elements for informed decision-making in materials science. Determining these associations—especially in the absence of ground truth values—demands a comprehensive understanding of data distribution, statistical relationships, and validation through methods such as p-values. A robust grasp of these aspects is vital for selecting appropriate statistical methodologies and ensuring accurate interpretations.

This paper also addresses the biases commonly encountered when evaluating feature importance derived from standard machine learning models, which can lead to erroneous conclusions. In the absence of accurate calculations in feature importances, this paper advocates for the use of multifaceted approaches using unsupervised machine learning models such as feature agglomeration and highly variable gene selection and followed by nonlinear nonparametric statistical methods such as Spearman's correlation. Three tools have a great feature ranking stability while PCA has feature ranking unstability due to linear nature of PCA.

2. Linear vs nonlinear models

Linear and nonlinear models represent fundamental approaches in machine learning for materials science. Linear models establish direct proportional relationships between input features and output targets, creating straightforward equations that make interpretation clear. They excel at predicting properties like tensile strength based on compositional data when relationships are relatively simple. Nonlinear models capture complex patterns through techniques like polynomial regression, decision trees, and neural networks, excelling at predicting behaviors involving complex interactions such as phase transitions. While linear models provide clear interpretability, nonlinear models can function as "black boxes" despite superior predictive power. Linear models require less computational power and resist overfitting with limited data, while nonlinear models demand more resources but

capture intricate behaviors common in materials science.

3. Parametric vs nonparametric

Parametric methods assume data follows a specific distribution and estimate parameters like mean and variance. Examples include linear regression, t-tests, and ANOVA, working well when relationships follow expected mathematical forms. Nonparametric methods make no distribution assumptions, focusing instead on data ranks or order. They prove valuable with small sample sizes or when parametric assumptions are violated, using techniques like the Wilcoxon rank-sum test. Parametric methods provide powerful tests when assumptions hold true, enabling clearer conclusions about material behavior. When these assumptions are violated, they yield misleading results. Nonparametric methods offer greater flexibility across diverse data types, though with potentially less statistical power.

4. True associations between variables

True associations represent actual connections between target variables and features. Understanding these relationships helps determine how feature changes impact outcomes, distinguishing between causation (direct impact) and correlation (variables moving together without causal links). Identifying true associations requires filtering out noise that obscures genuine relationships. True associations remain consistent across samples and experimental conditions, while effective preprocessing and statistical methods help isolate meaningful relationships. Genuine relationships typically align with underlying physical mechanisms and should demonstrate statistical significance and reproducibility across different datasets.

5. True relationships with machine learning models

Machine learning aims to predict targets with ground truth values for validation. However, feature importances lack ground truth values for accuracy validation, causing different models to generate different feature importances with inherent biases. Feature importance techniques (like SHAP) can inherit and amplify these biases. Feature selection methods like MLR-EM and LASSO identify contextually relevant features but rely on potentially biased importance assessments. Dimensionality reduction techniques like PCA help identify relevant features but may still be influenced by underlying biases. Without ground truth values for validation, feature importances derived from machine learning models remain inherently biased, potentially obscuring true relationships between variables and outcomes. Researchers should adopt robust validation techniques, employ multiple methodologies, and integrate domain expertise to mitigate these biases when interpreting feature importances.

6. Conclusion

In conclusion, this paper underscores the critical importance of a fundamental understanding of machine learning principles for researchers in materials science, essential for ensuring accurate analyses and reliable interpretations of results. By analyzing case studies that demonstrate the detrimental consequences of misapplying machine learning concepts, we emphasize the necessity of mastering both linear and nonlinear models, as well as parametric and nonparametric methods. The choice of these models profoundly impacts the reliability of findings related to material properties and behaviors, highlighting the urgent need for sound methodological frameworks.

Additionally, the ability to distinguish between correlation and causation is crucial for establishing authentic associations between target variables and their corresponding features in complex datasets. This task is often complicated by noise, which requires the use of robust statistical methods to clarify genuine relationships. As materials research increasingly incorporates advanced machine learning techniques, adopting error-free and bias-free methods becomes imperative to enhance the validity of outcomes.

As the field evolves, cultivating rigorous machine learning practices will be essential for improving the efficacy and reliability of materials analyses. By carefully selecting appropriate models and proactively addressing potential biases, researchers can significantly elevate their predictive capabilities. Collectively, these efforts will not only propel innovation in materials science but also contribute to the development of new materials with desirable attributes. This progress is poised to benefit a wide variety of industries and applications, showcasing the transformative potential of integrating robust statistical methods with machine learning in advancing materials research. Ultimately, the adoption of improved practices is vital for driving significant advancements that can redefine the future of materials science.

Authors' contributions

Yoshiyasu Takefuji completed this research and wrote the program and this article.

Ethics approval

Not applicable.

Consent to participate

Not applicable.

Consent for publication

Not applicable.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the author used GPT-4o mini in order to improve language and readability. After using this tool/service, the author reviewed and edited the content as needed and takes full responsibility for the content of the published article.

Funding

This research has no fund.

Conflicts of interest/Competing interests

The author has no conflict of interest.

Code availability

Not applicable

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.

References

[1] X. Zhao, Y. Yu, X. Xu, et al., Machine learning-assisted high-throughput screening of nanozymes for ulcerative colitis, Adv. Mater. (2025), 10.1002/adma.202417536.

[2] D.A. Winkler, A.E. Hughes, C. Özkan, et al., Impact of inhibition mechanisms, automation, and computational models on the discovery of organic corrosion inhibitors, Prog. Mater. Sci. 149 (2025) 101392, https://doi.org/10.1016/j.pmatsci.2024.101392.

[3] J. Hu, Y. Zou, Y. Deng, et al., Recent advances in non-ionic surfactant templated synthesis of porous metal oxide semiconductors for gas sensing applications, Prog. Mater. Sci. 150 (2025) 101409, https://doi.org/10.1016/j.pmatsci.2024.101409.

[4] M. Chen, K. Papadikis, C. Jun, N. Macdonald, Linear, nonlinear, parametric and nonparametric regression models for nonstationary flood frequency analysis, J. Hydrol. 616 (2023) 128772, https://doi.org/10.1016/j.jhydrol.2022.128772.

[5] R.J. Janse, T. Hoekstra, K.J. Jager, et al., Conducting correlation analysis: important limitations and pitfalls, Clin. Kidney J. 14 (11) (2021) 2332–2337, https://doi.org/10.1093/ckj/sfab085.

[6] K.R. Moon, D. van Dijk, Z. Wang, S. Gigante, D.B. Burkhardt, W.S. Chen, S. Krishnaswamy, Visualizing structure and transitions in high-dimensional biological data, Nat. Biotechnol. 37 (12) (2019) 1482–1492, https://doi.org/10.1038/s41587-019-0336-3.

[7] E. Matusik, O. Vassal, A. Conrad, et al., Parametric and nonparametric population pharmacokinetic analysis of fluconazole in critically ill patients and dosing simulations for candida infections, Antimicrob. Agents Chemother. 68 (11) (2024) e0099124, https://doi.org/10.1128/aac.00991-24.

[8] Politi M.T. Ferreira, J.C. Patino, CM. Nonparametric statistical tests: friend or foe? J. Bras. Pneumol. 47 (4) (2021) e20210292 https://doi.org/10.36416/1806-3756/e20210292.

[9] C. Merino-Soto, A. Juárez-García, G.S. Escudero, F. Toledano-Toledano, Parametric and nonparametric analysis of the internal structure of the psychosocial work processes questionnaire (PROPSIT) as applied to workers, Int. J. Environ. Res. Public Health 19 (13) (2022) 7970, https://doi.org/10.3390/ijerph19137970.

[10] GitHub. mnisttest.py. ⟨https://github.com/y-takefuji/mnist⟩.