Arthritis & Rheumatology
Vol. 0, No. 0, Month 2025, pp 1–2

American College of Rheumatology
Empowering Rheumatology Professionals

LETTER

DOI 10.1002/art.43222

Limitations of machine learning–based feature importance in osteoarthritis biomarker discovery: comment on the article by Kang et al

To the Editor:

This critical review examines methodologic concerns in the proteomic prediction model for osteoarthritis (OA) risk using Light Gradient-Boosting Machine (LightGBM) and SHapley Additive exPlanations (SHAP) by Kang et al.[1] Despite impressive predictive accuracy (area under the curve [AUC] > 0.8), we identify three fundamental limitations: (1) feature importance metrics lack ground-truth validation mechanisms, unlike prediction accuracy; (2) SHAP's model-dependent nature inherently incorporates and potentially amplifies model biases; and (3) high prediction accuracy does not guarantee reliable feature importance attribution, particularly in high-dimensional proteomic datasets. We propose alternative approaches using nonlinear nonparametric methods such as Mutual Information (MI) and Effective Transfer Entropy (ETE) for more reliable biomarker identification that better captures complex biologic relationships and reduces algorithmic artifacts in OA risk prediction.

Kang et al conducted a groundbreaking investigation using plasma proteomic profiles to predict individual OA risk.[1] Their predictive framework employed the LightGBM machine learning algorithm with SHAP for feature importance assessment. The model exhibited impressive predictive performance with AUC values reaching 0.820 for hip OA and 0.803 for knee OA in 5-year predictions.[1]

This paper, however, raises critical methodologic concerns regarding the implementation of LightGBM and SHAP techniques. Three fundamental limitations deserve particular attention. First, although the predictive accuracy of supervised machine learning models like LightGBM can be validated against ground-truth outcome values, the feature importance metrics derived from these models lack similar validation mechanisms, potentially leading to biased and distorted interpretations of biomarker significance.[2,3] Second, SHAP's model-dependent nature (functioning as explain = SHAP[model]) means it inherently incorporates and may even amplify the underlying model's biases in feature importance attribution, rather than providing model-independent validation.[4–6] Third, high target prediction accuracy does not guarantee reliable feature importances because of the absence of ground-truth values for feature contributions—a model might achieve excellent predictions while severely misattributing importance among features, particularly in high-dimensional proteomic datasets.[7–9]

Although this paper acknowledges the high prediction accuracy of LightGBM, it problematically relies on feature importances derived from LightGBM with SHAP without addressing these known limitations. The scientific literature contains more than 100 peer-reviewed articles documenting critical distortion issues in feature importance metrics derived from machine learning models. SHAP solely relies on the given model and inherits, and may potentially amplify, biases in feature importances derived from machine learning algorithms such as LightGBM, phenomena and limitations that have been extensively documented in peer-reviewed research. This substantial body of evidence suggests that proteomic biomarker identification based solely on such metrics should be approached with considerable caution, as potentially important markers might be overlooked while others could be falsely prioritized based on algorithmic artifacts rather than genuine biologic relevance to OA pathogenesis.

This paper advocates for alternative approaches using nonlinear nonparametric robust statistical methods such as MI analysis[10] and ETE for feature importance assessment. These methods offer distinct advantages for analyzing complex interactions among multiple variables with nonmonotonic patterns. Unlike model-dependent approaches, MI quantifies statistical dependencies between variables without assuming specific functional relationships, and ETE measures directed information flow between variables while controlling for confounding effects. These model-agnostic approaches provide more reliable feature importance metrics that are less susceptible to algorithmic biases and can better capture the complex, nonlinear relationships often present in biologic systems, potentially yielding more clinically relevant biomarkers for OA risk prediction.

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/art.43222.

Yoshiyasu Takefuji, PhD
takefuji@keio.jp
Musashino University
Tokyo, Japan

1. Kang Z, Zhang J, Liu W, et al. Plasma proteomic profiles predict individual future osteoarthritis risk. Arthritis Rheumatol Published online February 24, 2025. doi:https://doi.org/10.1002/art.43143
2. Nalenz M, Rodemann J, Augustin T. Learning de-biased regression trees and forests from complex samples. Mach Learn 2024;113(6):3379–3398.
3. Nazer LH, Zatarah R, Waldrip S, et al. Bias in artificial intelligence algorithms and recommendations for mitigation. PLOS Digit Health 2023;2(6):e0000278.

4. Bilodeau B, Jaques N, Koh PW, et al. Impossibility theorems for feature attribution. Proc Natl Acad Sci USA 2024;121(2):e2304406120.
5. Huang X, Marques-Silva J. On the failings of Shapley values for explainability. Int J Approx Reason 2024;171:109112.
6. Hooshyar D, Yang Y. Problems with SHAP and LIME in interpretable AI for education: a comparative study of post-hoc explanations and neural-symbolic rule extraction. IEEE Access 2024;12: 137472–137490.
7. Lenhof K, Eckhart L, Rolli LM, et al. Trust me if you can: a survey on reliability and interpretability of machine learning approaches for drug sensitivity prediction in cancer. Brief Bioinform 2024;25(5): bbae379.
8. Mandler H, Weigand B. A review and benchmark of feature importance methods for neural networks. ACM Comput Surv 2024; 56(12):318.
9. Potharlanka JL, Bhat M N. Feature importance feedback with Deep Q process in ensemble-based metaheuristic feature selection algorithms. Sci Rep 2024;14(1):2923.
10. Gibson JD. Entropy and mutual information. In: Gibson JD. Information Theoretic Principles for Agent Learning. Springer; 2025: 5–12.