Veterinary Microbiology 312 (2026) 110838

Contents lists available at ScienceDirect

Veterinary Microbiology

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

Letter to the Editor

From bias to reliable insight: Rethinking feature importance in microbiome analytics

A R T I C L E I N F O

Keywords:
Gut microbiota dysbiosis
Fowl adenovirus
Random forest
Feature importance
Model-agnostic methods
Interpretability

Kim et al. (2025) present an important experimental framework to investigate how antibiotic-induced gut microbiota dysbiosis influences antiviral immunity in chickens—a question of growing relevance in veterinary microbiology. Their study employed VNAM pre-treatment followed by FAdV-4 challenge, resulting in complete mortality by 4 dpi, compared to 54 % survival at day 8 in the infection-only group. Necropsy revealed severe hepatic lesions and liquefied cecal contents, while 16S rRNA sequencing demonstrated a pronounced Firmicutes/Bacteroidota ratio shift (CTRL: 90.0 % vs. 8.4 %; VNAM + FAdV: 94.0 % vs. 1.5 %) and depletion of SCFA-producing genera such as Alistipes and Butyricoccus. These findings underscore the biological significance of gut microbiota in antiviral defense and support computational approaches for identifying taxa associated with dysbiosis.

However, the interpretability of Random Forest-derived feature importance warrants careful scrutiny. Although this method is employed as a complementary tool, its application introduces critical concerns stemming from biases inherent in the analytical pipeline. Interpreting feature importance solely through predictive accuracy can be problematic—a concern widely documented in machine learning literature (Strobl et al., 2007; Touw et al., 2013; van Giffen et al., 2022). Tree-based models, including Random Forest, exhibit inherent biases favoring variables that enable early splits, particularly in high-dimensional, correlated datasets typical of microbiome research (Salles et al., 2021; Ishwaran, 2007). Consequently, taxa ranked as "important" may reflect statistical artifacts rather than mechanistic drivers, which could mislead biological inference if not rigorously validated.

Moreover, their study does not report hyperparameter settings, cross-validation strategy, or sensitivity analyses, which limits reproducibility. These omissions are critical because tuning parameters (e.g., number of trees, maximum depth) and validation protocols strongly influence feature rankings (Nalenz et al., 2023). Without transparency, it is impossible to determine whether the observed patterns are robust or artifacts of specific configurations. Post-hoc explanations, such as variable importance scores, inherit these biases and can vary with tuning choices, further reducing interpretability. This issue is particularly relevant in microbiome studies, where feature spaces are large and interdependent, making interpretability highly sensitive to algorithmic decisions.

The absence of external validation or replication across independent datasets is another critical limitation. Microbiome composition is highly context-dependent, and Random Forest models trained on small, homogeneous samples may overfit to local patterns rather than generalizable biological signals (Li et al., 2021). This risk is amplified by multicollinearity among microbial features, which destabilizes importance measures and inflates the apparent relevance of correlated taxa (Salles et al., 2021). Such instability underscores the need for alternative strategies that can disentangle correlation from causation and provide more robust interpretability.

To strengthen interpretability and improve their promising experimental design, we recommend a multifaceted analytical framework. First, unsupervised techniques such as Feature Agglomeration can mitigate multicollinearity and stabilize the feature space before classification (Zhang et al., 2020). Second, model-agnostic screening methods, analogous to Highly Variable Gene Selection, can identify informative taxa without embedding model-specific biases (Xie et al., 2025). Third, feature importance should be validated using non-parametric statistics, such as Spearman's rho or Kendall's tau, to confirm monotonic or non-linear associations (Okoye and Hosseini, 2024; Yu and Hutson, 2024). A feature should only be considered biologically meaningful if supported by both the explainer and statistical tests with significant p-values. Including such approaches would not only enhance interpretability but also align microbiome analytics with best practices in computational biology (Molnar, 2022).

Transparency in analytical pipelines is equally essential. Reporting hyperparameter settings, cross-validation strategies, and sensitivity analyses for feature importance would enhance reproducibility and allow readers to assess robustness. Without such transparency, even well-designed studies risk misinterpretation and hinder cumulative scientific progress. As microbiome research increasingly incorporates machine learning, adopting rigorous interpretability standards will be critical for translating computational findings into actionable biological insights and maintaining trust in computational approaches (Caminero et al., 2025).

https://doi.org/10.1016/j.vetmic.2025.110838
Received 29 November 2025; Accepted 10 December 2025

In conclusion, Kim et al.'s study provides a valuable experimental model and highlights the importance of gut microbiota in avian antiviral immunity. However, caution is needed when interpreting Random Forest-derived feature rankings as biologically meaningful indicators. Future work should integrate model-agnostic validation and diverse analytical frameworks to improve interpretability and translational potential in poultry disease research.

CRediT authorship contribution statement

Kiyo Yoshida: Investigation. **Yoshiyasu Takefuji:** Project administration, Supervision, Writing – review & editing. **Souichi Oka:** Conceptualization, Writing – original draft.

Funding Sources

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

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.

References

Caminero, A., Tropini, C., Valles-Colomer, M., et al., 2025. Credible inferences in microbiome research: ensuring rigour, reproducibility and relevance in the era of AI. Nat. Rev. Gastroenterol. Hepatol. https://doi.org/10.1038/s41575-025-01100-9.

Ishwaran, H., 2007. Variable importance in binary regression trees and forests. J. Am. Stat. Assoc. 102, 1184–1193. https://doi.org/10.1214/07-EJS039.

Kim, J.-H., Oh, B.S., Lee, R., Hunter, C.A., Cho, Y.-J., Park, J., 2025. Gut microbiota dysbiosis by antibiotics increases FAdV-4 susceptibility and alters antiviral immunity in chickens. Vet. Microbiol. 309, 110692. https://doi.org/10.1016/j.vetmic.2025.110692.

Li, Y., et al., 2021. Statistical and machine learning methods for microbiome data analysis. Front. Genet. 12, 682804. https://doi.org/10.3389/fgene.2021.682804.

Molnar, C., 2022. Interpretable Machine Learning. ⟨https://christophm.github.io/interpretable-ml-book/⟩.

Nalenz, M., Rodemann, J., Augustin, T., 2023. Learning de-biased regression trees and forests from complex samples. Mach. Learn. 113, 3379–3398. https://doi.org/10.1007/s10994-023-06439-1.

Okoye, K., Hosseini, S., 2024. Correlation tests in R: Pearson Cor, Kendall's Tau, and Spearman's Rho. R Programming: Statistical Data Analysis in Research. Springer, Singapore, pp. 247–277. https://doi.org/10.1007/978-981-97-3385-9_12.

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

Strobl, C., Boulesteix, A.L., Zeileis, A., Hothorn, T., 2007. Bias in random forest variable importance measures: Illustrations, sources and a solution. BMC Bioinform. 8, 25. https://doi.org/10.1186/1471-2105-8-25.

Touw, W.G., Bayjanov, J.R., Overmars, L., Backus, L., Boekhorst, J., Wels, M., van Hijum, S.A.F.T., 2013. Data mining in the life sciences with random forest: a walk in the park or lost in the jungle. Brief. Bioinform. 14, 315–326. https://doi.org/10.1093/bib/bbs034.

van Giffen, B., Herhausen, D., Fahse, T., 2022. Overcoming the pitfalls and perils of algorithms: A classification of machine learning biases and mitigation methods. J. Bus. Res. 144, 93–106. https://doi.org/10.1016/j.jbusres.2022.01.076.

Xie, Y., Jing, Z., Pan, H., Xu, X., Fang, Q., 2025. Redefining the high variable genes by optimized LOESS regression with positive ratio. BMC Bioinform. 26, 104. https://doi.org/10.1186/s12859-025-06112-5.

Yu, H., Hutson, A.D., 2024. A robust Spearman correlation coefficient permutation test. Commun. Stat. Theory Methods 53, 2141–2153. https://doi.org/10.1080/03610926.2022.2121144.

Zhang, J., Wu, X., Hoi, S.C.H., Zhu, J., 2020. Feature agglomeration networks for single stage face detection. Neurocomputing 380, 180–189. https://doi.org/10.1016/j.neucom.2019.10.087.

Souichi Oka*[,1], Kiyo Yoshida[2]
Science Park Corporation, 3-24-9 Iriya-Nishi, Zama-shi, Kanagawa 252-0029, Japan

Yoshiyasu Takefuji[3]
Faculty of Data Science, Musashino University, 3-3-3 Ariake Koto-ku, Tokyo 135-8181, Japan
E-mail address: takefuji@keio.jp.

* Corresponding author.
E-mail addresses: souichi.oka@sciencepark.co.jp (S. Oka), kyoshida@sciencepark.co.jp (K. Yoshida).

---

[1] ORCID: 0009-0000-4840-5232.

[2] ORCID: 0009-0007-3664-5547.

[3] ORCID: 0000-0002-1826-742X.