February 14, 2024

Prof. Paul Kononoff

Editor in Chief, Journal of Dairy Science

Dear Prof. Kononoff,

I am writing to express my sincere gratitude for the opportunity to submit a review paper to the Journal of Dairy Science (JDS). Although my initial submission was not accepted, I am appealing this decision, motivated by my commitment to enhancing the quality of machine learning (ML) research within the JDS community.

**Addressing the Key Concerns**

In the field of ML, reproducibility and comparability are paramount for any scientific study. However, a notable trend in recent JDS publications reveals a lack of detail necessary to meet these standards. This issue is compounded by the rise of tools like ChatGPT, which allow ML code generation with minimal understanding of underlying models. This presents a new challenge in research integrity, demanding extra efforts from reviewers and editors.

ML models are double-edged swords. They range from random forests to neural networks and offer powerful insights into non-linear relationships in data. Yet, their complexity can also lead to overfitting. The traditional statistical tests used in JDS papers, like ANOVA, cannot validate these complex models. This discrepancy underscores a significant gap in current ML research practices within the journal.

**Proposing Solutions**

I propose two critical measures to address these issues:

* Performance Metrics: Clearly defined metrics are essential for objective evaluation and comparison. For example, Confusion can arise from metric *R2* if it's not specified whether it refers to the squared Pearson correlation coefficient or the coefficient of determination. Clarifying this in each study is crucial to ensure everyone interprets the results similarly.
* Model Validation: Validation is crucial to simulate unseen data and ensure reproducibility. The portion of data withheld for validation impacts the performance metrics, a detail often overlooked but crucial for accurate comparison. For instance, can we consider the model performance to be equal between two studies if they both report an *R2* of 0.85 as their prediction accuracy, but one study withholds 10% of the data for validation while the other withholds 20%?

While these principles are discussed in various statistical literatures, there is a lack of a consolidated guide specifically for ML in dairy science. My review paper aims to fill this void by providing comprehensive guidelines supplemented by simulation experiments and practical examples from JDS publications.

**Case Studies and Improvements**

To provide more concrete examples, I have conducted a survey on six ML papers from the JDS. Five of them are top results by searching on Google Scholar with the keyword "journal of dairy science machine learning." The remaining one was chosen to represent the mentioned concern. In my analysis of the six papers, Becker et al. (2021) and Mota et al. (2021) stand out for their exemplary practices. Both studies effectively demonstrated model validation and clearly defined evaluation metrics, setting a standard for future studies. However, some papers did present issues:

* Ghaffari et al. (2019) (cited by 58): The model validation approach was not correctly implemented. Feature selection and hyperparameter tuning were omitted from the cross-validation process. Additionally, the risk of overfitting was heightened by the small sample size (38 cows) relative to the number of predictors (170 metabolites). This type of oversight is discussed in "The Elements of Statistical Learning" by Hastie et al. (2009, p. 247), which the review paper references.
* Frizzarin et al. (2021a) (cited by 49): On page 7440, the same dataset was improperly used for both hyperparameter tuning (the number of factors in partial least squares regression) and external validation, potentially biasing the model. Furthermore, on page 7442, the use of the term "cross-validation data" is inappropriate.
* Brand et al. (2021) (cited by 38): In Table 1 on page 4985, the authors did not clarify how hyperparameters, such as retention rate and random selection rate, were determined during cross-validation. Moreover, the method of external validation to confirm the model’s accuracy is not adequately described, with only a brief mention on page 4988 in the discussion section.
* Frizzarin et al. (2021b) (cited by 12): The study lacks any form of cross-validation. The method of selecting the validation set is neither random nor detailed, potentially skewing the results toward the chosen dataset.

**Proposed Actions for Revision**

Feedback on the initial submission highlighted two main areas for improvement: the abstract nature of the paper and the inclusion of unnecessary ML theory details. To address these, I plan to collaborate with Dr. Robin White, a renowned nutritionist with extensive experience in ML, to provide more relatable dairy science examples and streamline the theoretical content. This revision aims to make the paper more engaging and valuable to the JDS community. In conclusion, I respectfully request reconsideration of my submission. The revised paper will significantly contribute to the rigor and clarity of ML research in dairy science. I appreciate your consideration.

Sincerely,

  
**Chun-Peng James Chen, Ph.D.**  
Assistant Professor of Animal Data Sciences

Virginia Tech, Blacksburg, VA

**References**

Becker, C.A., A. Aghalari, M. Marufuzzaman, and A.E. Stone. 2021. Predicting dairy cattle heat stress using machine learning techniques. Journal of Dairy Science 104:501–524. doi:10.3168/jds.2020-18653.

Brand, W., A.T. Wells, S.L. Smith, S.J. Denholm, E. Wall, and M.P. Coffey. 2021. Predicting pregnancy status from mid-infrared spectroscopy in dairy cow milk using deep learning. Journal of Dairy Science 104:4980–4990. doi:10.3168/jds.2020-18367.

Frizzarin, M., I.C. Gormley, D.P. Berry, T.B. Murphy, A. Casa, A. Lynch, and S. McParland. 2021a. Predicting cow milk quality traits from routinely available milk spectra using statistical machine learning methods. Journal of Dairy Science 104:7438–7447. doi:10.3168/jds.2020-19576.

Frizzarin, M., T.F. O’Callaghan, T.B. Murphy, D. Hennessy, and A. Casa. 2021b. Application of machine-learning methods to milk mid-infrared spectra for discrimination of cow milk from pasture or total mixed ration diets. Journal of Dairy Science 104:12394–12402. doi:10.3168/jds.2021-20812.

Ghaffari, M.H., A. Jahanbekam, H. Sadri, K. Schuh, G. Dusel, C. Prehn, J. Adamski, C. Koch, and H. Sauerwein. 2019. Metabolomics meets machine learning: Longitudinal metabolite profiling in serum of normal versus overconditioned cows and pathway analysis. Journal of Dairy Science 102:11561–11585. doi:10.3168/jds.2019-17114.

Hastie, T., R. Tibshirani, and J.H. Friedman. 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer series in statistics. Springer.

Mota, L.F.M., S. Pegolo, T. Baba, F. Peñagaricano, G. Morota, G. Bittante, and A. Cecchinato. 2021. Evaluating the performance of machine learning methods and variable selection methods for predicting difficult-to-measure traits in Holstein dairy cattle using milk infrared spectral data. Journal of Dairy Science 104:8107–8121. doi:10.3168/jds.2020-19861.