Green Chemical Engineering xxx (xxxx) xxx

Contents lists available at ScienceDirect

Green Chemical Engineering

journal homepage: www.keaipublishing.com/en/journals/green-chemical-engineering

Critical evaluation of feature importance assessment in FFNN-based models for predicting Kamlet-Taft parameters

ARTICLE INFO

Keywords
Feature importance
Machine learning
FFNN
SHAP
Statistical significance
Spearman's correlation

ABSTRACT

Mohan et al. developed a feed-forward neural network (FFNN) model to predict Kamlet-Taft parameters using quantum chemically derived features, achieving notable predictive accuracy. However, this study raises concerns about conflating prediction accuracy with feature importance accuracy, as high R^2 and low RMSE do not guarantee valid feature importance assessments. The reliance on SHAP (SHapley Additive exPlanations) for feature evaluation is problematic due to model-specific biases that could misrepresent true associations. A broader understanding of data distribution, statistical relationships, and significance testing through p-values is essential to rectify this. This paper advocates for employing robust statistical methods, like Spearman's correlation, to effectively assess genuine associations and mitigate biases in feature importance analysis.

Mohan et al. [1] developed a feed-forward neural network (FFNN)-based machine learning (ML) model to predict the Kamlet-Taft parameters of designer solvents, employing quantum chemically derived input features. Their study demonstrated impressive predictive accuracy, as evidenced by high R^2 values and low root mean square error (RMSE) figures. To further investigate feature importance, the authors utilized SHAP (SHapley Additive exPlanations) analysis in conjunction with their FFNN model [1]. However, while acknowledging the high performance of the models proposed by Mohan et al., this paper raises critical concerns about the application of the FFNN model for extracting feature importances. A significant issue lies in their conflation of prediction accuracy with feature importance accuracy. Elevated R^2 values and low RMSE do not inherently translate to precise assessments of feature importance, which remains a distinct aspect that requires careful consideration.

Cross-validation, which focuses on manipulating data rather than models, is effective only for evaluating the accuracy of target prediction models; it does not yield reliable insights into feature importance accuracy. Consequently, high prediction accuracy does not guarantee accurate feature importance. Similarly, a high R^2 value and low RMSE do not ensure that the feature importance estimates are correct. While machine learning target predictions are accompanied by ground truth values that enable accuracy validation, feature importances do not have similar ground truth values. Consequently, cross-validation can effectively assess prediction accuracy but falls short when it comes to validating the accuracy of feature importance.

The reliance on SHAP for feature importance analysis raises additional issues, as SHAP values are inherently model-specific and can inherit and amplify biases from the underlying models, leading to erroneous conclusions [2–10]. Specifically, the FFNN model may consistently yield biased feature importance results due to its inability to adequately account for the data distribution and the statistical relationships between the target variables and the features. To derive genuine associations between the target and features, it is essential to incorporate three critical elements: a comprehensive understanding of data distribution, an examination of statistical relationships, and an assessment of statistical significance through p-values. Neglecting these aspects could result in misleading or inaccurate interpretations of feature importance derived from the FFNN model.

SHAP relies entirely on the provided model and inherently amplifies any biases present within it. To accurately capture the true associations or genuine relationships between the target and features, three key elements are essential: the data distribution, an examination of the relationships between the variables, and statistical validation through p-values. Selecting the appropriate modeling approach—whether a linear model or a nonlinear model with parametric or non-parametric methods—based on the data distribution is crucial for achieving bias-free computations. While machine learning target predictions possess ground truth values for prediction accuracy, feature importances lack in ground truth values. This paper advocates for integrating machine learning target predictions and Spearman's correlation with p-values, nonlinear and nonparametric approaches.

To enhance the rigor of their analysis, researchers like Mohan et al. must grasp several fundamental principles in machine learning: 1) machine learning models can consistently produce biased feature importance due to their model-specific nature, a concern that has been highlighted in over 100 peer-reviewed articles; 2) SHAP inherently inherits biases from the models upon which it is based; 3) high prediction accuracy does not guarantee accurate feature importance assessments; 4) model-specific biases in feature importance cannot be entirely mitigated; and 5) the processes of generating predictions and deriving feature importance are fundamentally different. This paper advocates for the assessment of true associations or genuine relationships between the target and features using bias-free robust statistical methods, such as Spearman's correlation coupled with p-value evaluations, to minimize bias and enhance validity. Spearman's correlation with p-values gives nonlinear and nonparametric approaches.

https://doi.org/10.1016/j.gce.2025.01.003
Received 3 December 2024; Received in revised form 10 December 2024; Accepted 9 January 2025

Please cite this article as: Y. Takefuji, Critical evaluation of feature importance assessment in FFNN-based models for predicting Kamlet-Taft parameters, Green Chemical Engineering, https://doi.org/10.1016/j.gce.2025.01.003

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

[1] M. Mohan, N. Gugulothu, S. Guggilam, R.T. Rajitha, M.K. Kidder, J.C. Smith, Physics-informed machine learning to predict solvatochromic parameters of designer solvents with case studies in CO_2 and lignin dissolution, Green Chem. Eng. (2024), https://doi.org/10.1016/j.gce.2024.11.003.

[2] B. Bilodeau, N. Jaques, P.W. Koh, B. Kim, Impossibility theorems for feature attribution, Proc. Natl. Acad. Sci. U. S. A. 121 (2024) e2304406120.

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

[4] R. Wang, P. Chaudhari, C. Davatzikos, Bias in machine learning models can be significantly mitigated by careful training: evidence from neuroimaging studies, Proc. Natl. Acad. Sci. U. S. A. 120 (2023) e2211613120.

[5] P.R. Bassi, S.S. Dertkigil, A. Cavalli, Improving deep neural network generalization and robustness to background bias *via* layer-wise relevance propagation optimization, Nature Commun. 15 (2024) 291.

[6] D. Güllmar, N. Jacobsen, A. Deistung, D. Timmann, S. Ropele, J.R. Reichenbach, Investigation of biases in convolutional neural networks for semantic segmentation using performance sensitivity analysis, Z. Med. Phys. 32 (2022) 346–360.

[7] A. Demircioğlu, Measuring the bias of incorrect application of feature selection when using cross-validation in radiomics, Insights Imaging 12 (2021) 172.

[8] C. Strobl, A.L. Boulesteix, A. Zeileis, T. Hothorn, Bias in random forest variable importance measures: illustrations, sources and a solution, BMC Bioinf. 8 (2007) 1–21.

[9] S. van de Geer, The bias of the lasso and worst possible sub-directions, in: J.-M. Morel, B. Teissier (Eds.), Lecture Notes in Mathematics, Springer, Cham, 2016, pp. 41–60.

[10] J. Krawczuk, T. Łukaszuk, The feature selection bias problem in relation to high-dimensional gene data, Artif. Intell. Med. 66 (2016) 63–71.

Yoshiyasu Takefuji*

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

* Corresponding author.

E-mail address: takefuji@keio.jp.