Strengthening Feature-Importance Interpretation in Pediatric Telemedicine Prediction Models

To the Editor:

Brintz et al developed eXtreme Gradient Boosting (XGBoost)-based predictive models to identify potentially "sick" children and cases requiring escalation within a pediatric telemedicine and medication delivery service in Haiti.[1] Their study represents an important proof-of-concept for integrating machine learning into electronic clinical decision support, particularly in resource-limited settings where early recognition of clinical deterioration is challenging. The reported discrimination performance, with an area under the receiver operating characteristic curve of 0.82 for moderate or severe disease and approximately 0.77–0.78 for escalation prediction, is promising. However, given the clinical implications of model-guided triage, the interpretation and stability of feature-importance rankings warrant further methodological discussion.

In this context, XGBoost gain-based importance may not directly correspond to clinically stable determinants of disease severity. Brintz et al calculated predictor importance using built-in XGBoost output based on gain in model performance and further used SHapley Additive exPlanations (SHAP) to assess the direction and magnitude of feature contributions. Importantly, high predictive accuracy does not necessarily validate the clinical reliability of feature-importance rankings.[2] In telemedicine-derived datasets, symptom-based predictors may be correlated, incompletely observed, and shaped by parental reporting, provider interpretation, and branching logic in case report forms. Under these conditions, tree-based importance metrics may emphasize variables that improve decision-tree splits rather than robust clinical signals.[3] SHAP-based explanations provide an additional interpretive layer but remain model-dependent and sensitive to background distributions, missingness patterns, and correlated predictors.[4] Therefore, SHAP values should not be interpreted as direct evidence of causal or clinically fundamental importance without further validation, particularly when feature rankings inform clinical decision-support design.

To strengthen interpretability, we advocate a complementary feature-interpretation framework incorporating unsupervised feature-selection and feature-structuring approaches. Highly variable gene selection–inspired variance-based feature selection and feature agglomeration may help reduce redundancy, mitigate multicollinearity, and stabilize feature structure before supervised modeling. These steps could be complemented by model-agnostic non-parametric methods such as Spearman's rho and Kendall's tau to evaluate monotonic associations and ranking stability across resampling, model refitting, and clinical subgroups.[5] Such analyses would help determine whether observed feature attributions are reproducible and clinically meaningful, rather than artifacts of limited sample size, missingness, or correlated telemedicine features.

In summary, Brintz et al provide an important step toward AI-supported pediatric telemedicine. Complementing their framework with feature-structure diagnostics and model-agnostic validation would help ensure that XGBoost-based clinical decision support yields interpretable, reproducible, and clinically meaningful insights.

CRediT authorship contribution statement

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

Declaration of Competing Interests

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.

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

Souichi Oka, PhD
Faculty of Data Science
Musashino University
Koto-ku, Tokyo, Japan

Seiji Saito, PhD
Manufacturing Technology Department
The Polytechnic University of Japan
Kodaira-shi, Tokyo, Japan

Yoshiyasu Takefuji, PhD
AI Research Institute
Science Park Corporation
Zama-shi, Kanagawa, Japan

https://doi.org/10.1016/j.jpeds.2026.115268

References

1. Brintz BJ, Klarman MB, Cajusma Y, Exantus L, Beausejour JR, Flaherty KE, et al. Predictive modeling of childhood disease severity at a telemedicine service: evidence from the Improving Nighttime Access to Care and Treatment studies in Haiti. J Pediatr 2026;293:115032. https://doi.org/10.1016/j.jpeds.2026.115032

2. Musolf AM, Holzinger ER, Malley JD, Bailey-Wilson JE. What makes a good prediction? Feature importance and beginning to open the black box of machine learning in genetics. Hum Genet 2022;141:1515-28. https://doi.org/10.1007/s00439-021-02402-z
3. Adler AI, Painsky A. Feature importance in gradient boosting trees with cross-validation feature selection. Entropy 2022;24:687. https://doi.org/10.3390/e24050687
4. Bilodeau B, Jaques N, Koh PW, Kim B. Impossibility theorems for feature attribution. Proc Natl Acad Sci U S A 2024;121:e2304406120. https://doi.org/10.1073/pnas.2304406120
5. Okoye K, Hosseini S. Correlation tests in R: pearson cor, Kendall's tau, and Spearman's rho. In: Okoye K, Hosseini S, eds. R programming: statistical data analysis in research. Singapore: Springer Nature; 2024. p. 247-77.