Intensive Care Med (2023) 49:1559
https://doi.org/10.1007/s00134-023-07252-z

CORRESPONDENCE

SHAP model explainability in ECMO–PAL mortality prediction: a critical analysis

Marcos Valiente Fernández*, Amanda Lesmes González de Aledo, Francisco de Paula Delgado Moya and Isaías Martín Badía

We have read with great interest the work by Stephens et al. [1].

Data science not only comprises the use of complex statistical tools, but also the use of graphics and plots to display the results.

In critical care, understanding machine learning "black boxes" is increasingly important. Many methods have emerged to offer both local and global explanations of these algorithms [2]. Local explanations, like the LIME (Local Interpretable Model-agnostic Explanations) method, address specific predictions, asking 'Why did patient X experience this outcome?'. Conversely, the SHAP (SHapley Additive exPlanations) methodology, based on game theory, provides a broader view, illustrating how each data feature affects overall predictions.

In this article, SHAP model explainability is employed, and its results are presented using a plot.

This plot, though complex, details both the importance and the effect of the variables:

- In the algorithm decision model, feature importance is ordered hierarchically from top to bottom, with lactate and age being the most important.
- X-axis represents the Shapley value which, in their words, reflects "the impact on the model's results". It shows whether the effect of that value is associated with a higher or lower prediction.
- On the X-axis, positive values indicate a positive correlation, and negative values indicate a negative correlation.
- "Original value" is represented by color. It shows whether that variable is high (in red) or low (in blue) for that observation.

However, when examining the figure, we notice a discrepancy. If we focus on lactate values, the positive ones on the X-axis (those on the right side), are associated with the variable of interest (mortality), but in the blue color; in other words, low values. Just the opposite of what is being described in the results, and contrary to what we would expect based on scientific evidence.

We suspect this may be due to a graphic error and recommend it be corrected for clarity and coherence.

Funding
None.

Declarations

Conflicts of interest
Authors declare no conflicts of interest.

Publisher's Note

Accepted: 10 October 2023
Published: 31 October 2023

*Correspondence: mvalientefernandez@gmail.com
Intensive Care Medicine Department, Hospital Universitario 12 de Octubre, Avda. de Córdoba, s/n, 28041 Madrid, Spain

This comment refers to the article available online at https://doi.org/10.1007/s00134-023-07157-x.

References

1. Stephens AF, Šeman M, Diehl A et al (2023) ECMO PAL: using deep neural networks for survival prediction in venoarterial extracorporeal membrane oxygenation. Intensive Care Med. https://doi.org/10.1007/s00134-023-07157-x
2. Linardatos P, Papastefanopoulos V, Kotsiantis S (2020) Explainable AI: a review of machine learning interpretability methods. Entropy (Basel) 23(1):18. https://doi.org/10.3390/e23010018