A Hybrid Modeling Framework for Generalizable and Interpretable Predictions of ICU Mortality: Leveraging ICD Codes in a Multi-Hospital Study of Mechanically Ventilated Influenza Patients

Moein E. Samadi[1,*], Jorge Guzman-Maldonado[1], Kateryna Nikulina[1], Hedieh Mirzaieazar[1], Konstantin Sharafutdinov[1], Sebastian Johannes Fritsch[2,3], and Andreas Schuppert[1]

[1]Institute for Computational Biomedicine, RWTH Aachen University, Aachen, Germany
[2]Department of Intensive Care Medicine, University Hospital RWTH Aachen, Aachen, Germany
[3]Jülich Supercomputing Centre, Forschungszentrum Jülich, Jülich, Germany
[*]moein.samadi@rwth-aachen.de

ABSTRACT

The development of reliable mortality risk stratification models is an active research area in computational healthcare. Mortality risk stratification provides a standard to assist physicians in evaluating a patient's condition or prognosis objectively. Particular interest lies in methods that are transparent to clinical interpretation and that retain predictive power when validated across diverse data sets they were not trained on.
We've developed a hybrid model integrating mechanistic, clinical knowledge with mathematical and machine learning models to predict ICU mortality using ICD codes. A tree-structured network connecting independent modules that carry clinical meaning is implemented for interpretability. The trained model is then validated on external data sets from different hospitals, demonstrating successful generalization capabilities.

Supplementary Information

The generalizability of a clinical prediction model is commonly tested by how well it works with new patient data, evaluated through an external validation study. However, the term "external validation" is loosely defined and can leave ambiguity regarding how different the derivation sample is from the validation sample[1]. To quantify to what extent the derivation and validation samples are related, we used the Jaccard similarity that measures the case mix between the derivation and validation samples.

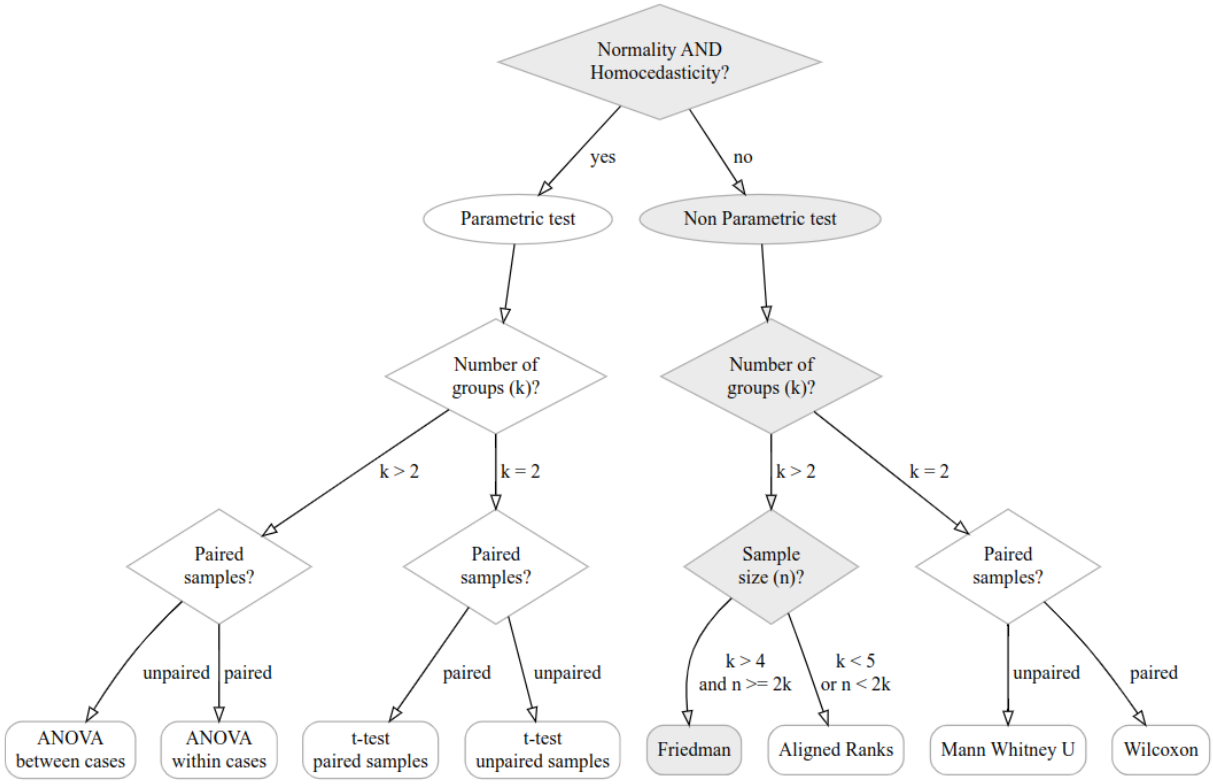
Jaccard similarity provides a simple, yet effective approach that determines the degree of relatedness between two sets of data. Given two sets A (derivation sample) and B (validation sample), the Jaccard similarity index J is calculated as:

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} \tag{1}$$

Where:

- $|A \cap B|$ is the count of shared attributes between the derivation and validation samples (intersection),
- $|A \cup B|$ is the count of all attributes present in either the derivation or validation samples, or both (union).

The value of $J(A,B)$ will range from 0 to 1, with 0 indicating no overlap or similarity, and 1 indicating complete overlap or similarity between the samples.



Supplementary Figure S1. The decision process for selecting the appropriate statistical test to quantitatively measure the consistency of the interpretations provided by SHAP values. The STAC library[2] was utilized to identify the optimal test. Due to the lack of normality and homoscedasticity in the SHAP value distributions, a non-parametric test was chosen. The Friedman test was identified as an appropriate choice for our analysis due to the necessity of comparing more than two distributions ($k = 5$). This test was particularly fitting, given that we utilized $n = 80$ ($\geq 2k$) validation data subsets in our examination.

	Derivation Hospital	Validation Hospital 1	Validation Hospital 2	Validation Hospital 3	Validation Hospital 4
Colsample by tree	0.5	0.75	0.75	0.75	0.5
Gamma	0	0.25	1	1	0
Learning rate	0.05	0.01	0.1	0.1	0.05
Max depth	4	4	4	4	4
Reg lambda	0	1	1	1	1
Subsample	0.75	0.75	0.75	0.5	0.5

Supplementary Table S1. Optimized hyperparameters in grid cross-validation for the training of the XGBoost model.

References

1. Debray, T. P. *et al.* A new framework to enhance the interpretation of external validation studies of clinical prediction models. *J. clinical epidemiology* **68**, 279–289 (2015).

2. Rodríguez-Fdez, I., Canosa, A., Mucientes, M. & Bugarín, A. STAC: a web platform for the comparison of algorithms using statistical tests. In *Proceedings of the 2015 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)* (2015).