

MATTERS ARISING

Open Access



A prediction model for venous thromboembolism in the intensive care unit: flawed methods may lead to inaccurate predictions

Stephen Gerry^{1*} and Gary S. Collins¹

In their recent article, Guan and colleagues describe the development of a prediction model for venous thromboembolism (VTE) using a large intensive care unit (ICU) multicentre data set from the USA [1]. We appreciate their efforts to produce a machine learning model that is not a typical ‘black-box’, but one where the model output is ‘interpretable’.

However, there are a few points on which we have some concerns. The first point is related to how the model estimates a patient’s predicted risk and whether these estimates are well calibrated. It is important that the output of a prediction model is a probability and not simply a classification (e.g. high risk vs low risk), since probabilities are so much more informative [2]. When a probability is presented to the end user, they are able to apply their own decision threshold. The model presented by Guan and colleagues does appear to produce predicted probabilities, as shown in Fig. 4; however, it is not clear how these probabilities are generated. Predicted probabilities are not typically produced by a random forest model, and therefore, a further stage of analysis is

normally necessary. Yet, this is not described. Since the model does appear to output probabilities to the user, it is important that the calibration of the model be assessed in the validation cohort. Calibration is a widely recommended performance measure and recommended in the TRIPOD (Transparent Reporting of a multivariable prediction model of Individual Prognosis Or Diagnosis) reporting guideline [3] and refers to the agreement between a model’s predicted risks and the observed risks. There are several ways to examine calibration; perhaps, the most effective is the calibration plot [4]. However, this paper does not include any assessment of model calibration. The calibration of a model can be impacted by the ‘overfitting’ inherent in the model fitting process, and random forest models are particularly prone to this [5] and are therefore more susceptible to miscalibration. Unless these issues are addressed, it is uncertain whether the risks generated by the model will be accurate or generalizable.

A second point on which we have concern is regarding missing data and the methods that were used to account for it during model development and validation. The authors helpfully describe the proportion of missing data for each variable in Supplementary Fig. 1, and in some cases, the proportion is very high, for example, 45.8% for partial thromboplastin time (PTT). The methods section states that multiple imputation was used to impute missing values. However, it is not clear what method was used to estimate the final model based upon these multiple imputed data sets during model development

This comment refers to the article available online at <https://doi.org/10.1186/s13054-023-04683-4>.

*Correspondence:

Stephen Gerry

Stephen.gerry@csm.ox.ac.uk

¹ Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Botnar Research Centre, Windmill Road, Oxford OX3 7LD, UK



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

or to estimate performance statistics during validation. Furthermore, other important information on the imputation approach is missing, such as the number of imputations, the imputation model and whether the outcome was included, and the rationale for assuming that imputation was appropriate [6, 7]. These issues are particularly relevant since the variable with the most missing data, PTT (45.8%), is also the variable with the greatest feature importance. Therefore, a poorly specified imputation model may have a considerable effect on how well the model works, and again, this calls into question the generalizability of the model.

Many of these omissions would likely have been addressed prior to publication had the authors used the TRIPOD reporting guideline, which is a tool to improve the reporting standard of clinical prediction (or diagnostic) models [3]. TRIPOD can already easily be used for artificial intelligence or machine learning models; however, there will soon be an updated version that more explicitly addresses factors that are unique to these types of models [3, 8].

Abbreviations

VTE	Venous thromboembolism
ICU	Intensive care unit
TRIPOD	Transparent Reporting of a multivariable prediction model of Individual Prognosis Or Diagnosis
PTT	Partial thromboplastin time

Acknowledgements

Not applicable.

Author contributions

SG drafted and GSC revised the manuscript.

Funding

No funding was associated with this manuscript.

Availability of data and materials

Not applicable.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 8 December 2023 Accepted: 12 December 2023

Published online: 18 December 2023

References

1. Guan C, Ma F, Chang S, Zhang J. Interpretable machine learning models for predicting venous thromboembolism in the intensive care unit: an analysis based on data from 207 centers. *Crit Care*. 2023;27(1):406.
2. Steyerberg EW, Moons KGM, van der Windt DA, Hayden JA, Perel P, Schroter S, et al. Prognosis research strategy (PROGRESS) 3: prognostic model research. *PLoS Med*. 2013;10(2):e1001381.
3. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med*. 2015;162(1):211–9.
4. Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW, Van Calster B, et al. Calibration: the Achilles heel of predictive analytics. *BMC Med*. 2019;17:1.
5. van der Ploeg T, Austin PC, Steyerberg EW, van der Ploeg T, Austin PC, Steyerberg EW. Modern modelling techniques are data hungry: a simulation study for predicting dichotomous endpoints. *BMC Med Res Methodol*. 2014;14(1):1–13.
6. Wood AM, Royston P, White IR. The estimation and use of predictions for the assessment of model performance using large samples with multiply imputed data. *Biometr J Biometr Z*. 2015;57(4):614–32.
7. Moons KGM, Donders RART, StijnenHarrell Jr TFE. Using the outcome for imputation of missing predictor values was preferred. *J Clin Epidemiol*. 2006;59(10):1092–101.
8. Collins GS, Dhiman P, Navarro CLA, Ma J, Hooft L, Reitsma JB, et al. Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence. *BMJ Open*. 2021;11(7):e048008.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.