

LETTER TO THE EDITOR

Open Access



Comment on: Machine Learning for Understanding and Predicting Injuries in Football

Garrett S. Bullock^{1,2,3*} , Patrick Ward⁴, Gary S. Collins⁵, Tom Hughes^{6,7} and Franco Impellizzeri⁸

Dear Editor,

We recently read the article titled “Machine Learning for Understanding and Predicting Injuries in Football” in *Sports Medicine – Open* [1]. Given that injury prediction is an emerging topic within sport, the increasing interest and excitement towards complex machine learning algorithms within this space is a cause for concern when fundamental principles of prediction model development are not followed. As such, we feel the need to intervene and highlight methodological and conceptual inaccuracies.

The models presented in this paper were deemed by the authors to be “quite sound” [1]. However, this is not true, as recently highlighted in the systematic review in *Sports Medicine* [2]. All of these models were included in this systematic review, and after evaluation with the

established Prediction Model Risk of Bias Assessment Tool (PROBAST) [3], were rated as high or unclear risk of bias [2].

The authors detail that, “the use of machine learning has great potential to unearth new insights into the workload and injury relationship” [1]. Prediction models may use both causal and non-causal predictors to estimate the risk of a future outcome [4, 5]. Consequently, it is inappropriate to use the included predictors to infer causal relationships between individual predictors and the outcome [6, 7]. Further, the authors state that Shapley values, local interpretable model-agnostic explanations, and partial dependency plots can be used to assist in interpreting cause-effect relationships with machine learning models [8]. These tools assess for associations between predictors and outcomes and, regardless of how these methods are labelled, the popular adage “correlation is not causation” still holds [8]. Importantly, these methods remain explorative, provide post hoc explanations (rationalization), and require confirmatory studies.

Such inaccurate and incorrect interpretations of clinical prediction models are of particular concern. This is because they can lead practitioners to attempt to change injury risk by intervening or manipulating predictor variables under the false assumption of a causal relationship; while these strategies are likely ineffective, they also have potentially harmful consequences for the athlete [4, 5, 9].

While the authors promote balancing dataset outcomes through over and under-sampling [1], this is highly discouraged as ‘balancing’ datasets alters the outcome prevalence, biasing towards overestimating risk. [10, 11] Balancing data without appropriate recalibration can inappropriately impact risk prediction

This comment refers to the article available online at <https://doi.org/10.1186/s40798-022-00465-4>.

*Correspondence:

Garrett S. Bullock
garretbullock@gmail.com

¹ Department of Orthopaedic Surgery and Rehabilitation, Wake Forest University School of Medicine, Winston-Salem, NC, USA

² Department of Biostatistics and Data Science, Wake Forest School of Medicine, Winston-Salem, NC, USA

³ Centre for Sport, Exercise and Osteoarthritis Research Versus Arthritis, University of Oxford, Oxford, UK

⁴ Seattle Seahawks, Seattle, WA, USA

⁵ Centre for Statistics in Medicine, University of Oxford, Oxford, UK

⁶ Department of Health Professions Institute of Sport, Manchester Metropolitan University, Manchester, UK

⁷ Institute of Sport, Manchester Metropolitan University, Manchester, UK

⁸ School of Sport, Exercise, and Rehabilitation, University of Technology Sydney, Sydney, Australia

and ultimately decision-making [11]. The authors also encourage creating classification models. Classification models are not recommended as this supersedes clinical and performance decision-making from the model users [11]. Classification models do not allow situational context and assume all situations and individuals have the same risk threshold. Prediction models should be developed and reported as a probability or at least risk score, to allow user interpretation and decisions [11].

The authors report that area under the curve (a form of discrimination), accuracy, sensitivity, and specificity should be used to evaluate machine learning models. These metrics are only a few of the recommended performance values that should be transparently reported [12] and some (such as accuracy) have well-known problems (also in machine learning) [13, 14]. The authors do not mention calibration as a model performance metric, which is the agreement between predicted and actual probabilities [15]. Calibration is always recommended and allows users to evaluate model performance across a range of risks [4, 5, 12]. Calibration divergence at specific risk ranges can alter clinical and performance decisions, even in models with high area under the curve performance [15].

The authors mention limitations of the included research, often only evaluating a single season of injuries for a given team, and how future research should seek to evaluate the models on subsequent seasons of data. Such a statement calls into question the notion of internal and external validation of these included papers. While the models in these papers have only been internally validated, their results provide a high level of optimism but no real understanding of how they may function under different circumstances or timepoints [16, 17]. Without a process of external validation for such models it is impossible to know whether or not they will be useful to practitioners with different clubs, observing different training constraints, under different coaches, and with different athletes who may have different base levels of risk [6, 18].

The purpose of this letter is to provide clarity about an important and burgeoning sports medicine and performance field. Other points that are not examined due to word count include sample size calculations, handling of missing data, data leakage, internal validation over train test splits, and clinical utility through clinical decision analysis and impact studies. Using the highest quality methods with transparency is imperative to protect and improve athlete health and performance, and to reduce the proliferation of conceptual and methodological inaccuracies about prediction model development and interpretation.

Acknowledgements

None.

Author contributions

GB, PW, and FM conceived the manuscript idea. GB, PW, and FM wrote the first draft of the manuscript. GB, PW, GC, TH, and FM critically revised the manuscript. GB, PW, GC, TH, and FM approved the final version of the manuscript.

Funding

No funding was obtained for this letter.

Availability of Data and Material

Data sharing is not applicable to this article as no datasets were generated or analysed during the current study.

Declarations

Ethics Approval and Consent to Participate

No ethics or consent approval were necessary for this letter.

Consent for Publication

Not applicable.

Competing interests

The authors declare that they have no competing interests with the content of this letter.

Received: 23 January 2024 Accepted: 25 June 2024

Published online: 28 July 2024

References

- Majumdar A, Bakirov R, Hodges D, Scott S, Rees T. Machine learning for understanding and predicting injuries in football. *Sports Med Open*. 2022;8:73.
- Bullock GS, Mylott J, Hughes T, Nicholson KF, Riley RD, Collins GS. Just how confident can we be in predicting sports injuries? A systematic review of the methodological conduct and performance of existing musculoskeletal injury prediction models in sport. *Sport Med*. 2022;52(10):2469–82.
- Wolff RF, Moons KG, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. *Ann Int Med*. 2019;170(1):51–8.
- Bullock GS, Hughes T, Sergeant JC, Callaghan MJ, Riley R, Collins G. Methods matter: clinical prediction models will benefit sports medicine practice, but only if they are properly developed and validated. *British J Sports Med*. 2021;55(23):1319–21.
- Bullock GS, Hughes T, Sergeant JC, Callaghan MJ, Riley RD, Collins GS. Clinical prediction models in sports medicine: a guide for clinicians and researchers. *J Orthop Sport Phys Ther*. 2021;51(10):517–25.
- Bullock GS, Hughes T, Arundale AH, Ward P, Collins GS, Kluzek S. Black box prediction methods in sports medicine deserve a red card for reckless practice: a change of tactics is needed to advance athlete care. *Sports Med*. 2022;52(8):1729–35.
- Riley RD, Hayden JA, Steyerberg EW, Moons KG, Abrams K, Kyzas PA, et al. Prognosis Research Strategy (PROGRESS) 2: prognostic factor research. *PLoS Med*. 2013;10(2): e1001380.
- Heskes T, Sijben E, Bucur IG, Claassen T. Causal shapley values: exploiting causal knowledge to explain individual predictions of complex models. *Adv Neural Inf Process Syst*. 2020;33:4778–89.
- Impellizzeri FM, Tenan MS, Kempton T, Novak A, Coutts AJ. Acute: chronic workload ratio: conceptual issues and fundamental pitfalls. *Int J Sport Phys Perform*. 2020;15(6):907–13.
- Harrell FE. *Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis*: Springer; 2001.
- van den Goorbergh R, van Smeden M, Timmerman D, Van Calster B. The harm of class imbalance corrections for risk prediction models: illustration and simulation using logistic regression. *J Am Med Inform Assoc*. 2022;29(9):1525–34.

12. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, et al. Assessing the performance of prediction models: a framework for some traditional and novel measures. *Epidemiol.* 2010;21(1):128.
13. Dinga R, Penninx BW, Veltman DJ, Schmaal L, Marquand AF. Beyond accuracy: measures for assessing machine learning models, pitfalls and guidelines. *BioRxiv.* 2019;8:743138.
14. Collins GS, Dhiman P, Navarro CL, Ma J, Hooft L, Reitsma JB, Logullo P, Beam AL, Peng L, Van Calster B, van Smeden M. 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.
15. Van Calster B, McLernon DJ, Van Smeden M, Wynants L, Steyerberg EW. Calibration: the Achilles heel of predictive analytics. *BMC Med.* 2019;17(1):1–7.
16. Collins GS, Altman DG. An independent external validation and evaluation of QRISK cardiovascular risk prediction: a prospective open cohort study. *BMJ.* 2009;7(339):b2584.
17. Collins GS, de Groot JA, Dutton S, Omar O, Shanyinde M, Tajar A, et al. External validation of multivariable prediction models: a systematic review of methodological conduct and reporting. *BMC Med Res Methodol.* 2014;14(1):1–11.
18. Bullock GS, Ward P, Impellizzeri FM, Kluzek S, Hughes T, Dhiman P, et al. The trade secret taboo: open science methods are required to improve prediction models in sports medicine and performance. *Sport Med.* 2023;53:1–9.

Publisher's Note

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