LETTER

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

Aritra Majumdar[1*] , Rashid Bakirov[1] and Tim Rees[2]

Dear Editor,

We acknowledge the Letter to the Editor by Bullock and colleagues [1] regarding our article "Machine learning for understanding and predicting injuries in football" [2], and appreciate the opportunity to respond. In our Leading Article [2], we outlined the topics of sport injury and machine learning, before describing examples from the literature that had used machine learning to examine the workload-injury relationship in football. Our aim was to "aid readers both from sport science and machine learning communities in their understanding of sports injury articles employing machine learning" (p. 2) [2]. We concluded: "the myriad ways machine learning can be employed can also lead to difficulty in synthesising the current research evidence into an overall, unified, conclusion. Indeed, there remain questions as to the utility of these models for real-world application" (p. 8) [2].

Given the above, we were confused as to the content and purpose of the letter [1]. The letter either (a) raised points with which we have not disagreed, backed by citations to the letter authors' own work, (b) made general observations about machine learning, or (c) countered points we never made—strawman logical fallacies.

The letter made five key points [1]. The first point—that we claimed the models that we reviewed in our Leading Article [2] were "quite sound"—is untrue. The letter authors [1] noted that the machine learning studies we reviewed were rated by them in a previous article [3] as having a "high or unclear risk of bias". We made it quite clear [2] that the machine learning *techniques* employed were legitimate, but that "there is considerable variability in study design and analysis" (p. 7), and "greater detail regarding the machine learning approaches employed would help any objective assessment of their contribution towards better understanding the workload–injury relationship" (p. 7). Thus, through our Leading Article [1]—not a systematic review—we arrived at roughly the same conclusion. Indeed, we wrote at length of our criticisms of the studies we reviewed, which included a lack of detail regarding the machine learning processes followed, such that they "limit a systematic evaluation of findings and the drawing of a unified conclusion" (p. 1) [2].

The second point—that we claimed that machine learning studies demonstrate causality—is also untrue. Nowhere in the article did we discuss causality or try to unpick cause-effect. We consistently used the term workload-injury *relationship*. One wonders, then, why the letter authors felt compelled to deliver such a clichéd, stats-class admonition; for further discussions on this topic, see [4–6].

The third point—that we promoted data balancing and classification models—is similarly untrue. Rather than

*Correspondence:
Aritra Majumdar
amajumdar@bournemouth.ac.uk
[1] Department of Rehabilitation and Sport Science, Faculty of Health and Social Sciences, Bournemouth University, Dorset House, Talbot Campus, Fern Barrow, Poole BH12 5BB, UK
[2] Department of Computing and Informatics, Faculty of Science and Technology, Bournemouth University, Dorset House, Talbot Campus, Fern Barrow, Poole BH12 5BB, UK

Majumdar *et al. Sports Medicine - Open*   (2024) 10:85

Page 2 of 3

"promoting", we noted that data balancing is used in machine learning, describing in the highlighted papers where data balancing had been employed [2]. We would agree that data balancing can have unintended consequences. Over-detection of medical conditions and subsequent over-treatment is one obvious example. The above notwithstanding, data balancing methods may still be critical in such contexts; for example, two recent articles demonstrated that data balancing improved the performance of machine learning models in diagnosing patients with metabolic dysfunction-associated steatohepatitis [7] and cerebrovascular diseases [8]. We also noted (rather than promoted) that the papers highlighted in our leading article had used classification models. Despite the letter's criticisms, many classification models also provide options to examine probability and risk scores [9].

The fourth point—that we advocated the use of some evaluation metrics at the expense of others—is also untrue. We did not state that certain metrics "should be used" [2]. We noted that many evaluation metrics are used, highlighting those metrics in the papers we reviewed, explaining their meaning, purpose, and limitations. It is true that we did not explicitly highlight "calibration"—something not mentioned in the papers we reviewed. More to the point, part of our Leading Article's [2] purpose was to "highlight (and to an extent de-mystify) the machine learning process" (p. 8) for non-experts. It was not designed as a treatise on machine learning. The argument that we did not mention calibration could be extended to wondering why we did not mention a host of other potentially important aspects of evaluation [10], such as the confusion matrix, Matthews correlation coefficient, and threat score.

We agree with the letter's fifth point—that the papers we reviewed in our Leading Article did not use external validation [2]. It is true that without "external validation for such models it is impossible to know whether or not they will be useful to practitioners with different clubs". It is a moot point, however, whether external validation is always necessary, and whether a universally applicable model for all clubs is desirable or even a logical endpoint of such research. In fact, it is precisely because different clubs use different training regimes, under different coaches, with different athletes, that a universal model would likely prove to be of little practical utility across all clubs. For example, a club may choose to play a distinctive, high tempo pressing style of football or a more strategic, defensive approach, potentially coupled with higher-than-average training volumes. To accurately predict injuries in such specific settings may mean developing predictive models that would not translate well to clubs with different styles of playing and training.

Relatedly, we recently reproduced with our own data [11] the analysis strategies from two of the papers [12, 13] highlighted in our Leading Article, observing sizeable discrepancies in the best models.

Finally, the letter noted, in passing, other issues that were not further expanded upon [1]. We shall thus not address these here, but again refer readers to our point that our article was not a treatise on machine learning.

Given the above, as a target for criticism, then, the choice of our Leading Article [2] seems strange. Although we welcome constructive peer review and the opportunity to provide clarifications of our work, we stress the importance of paying attention to the specifics of the article content, and making factual and unbiased observations of others' work.

Declarations

Competing interests
Aritra Majumdar, Rashid Bakirov and Tim Rees declare that they have no conflicts of interest relevant to the content of this letter.

Published online: 29 July 2024

References

1. Bullock GS, Ward P, Collins GS, Hughes T, Impellizzeri FM. Comment on: "machine learning for understanding and predicting injuries in football." Sports Med Open. 2022;8(1):73.
2. Majumdar A, Bakirov R, Hodges D, et al. Machine learning for understanding and predicting injuries in football. Sports Med-Open. 2022;8:73.
3. Bullock GS, Mylott J, Hughes T, et al. 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. Sports Med. 2022;52(10):2469–82.
4. Hernán MA. The C-word: scientific euphemisms do not improve causal inference from observational data. Am J Public Health. 2018;108(5):616–9.
5. Dahabreh IJ, Bibbins-Domingo K. Causal inference about the effects of interventions from observational studies in medical journals. JAMA. 1845;331(21):2024.
6. Carloni G, Berti A, Colantonio S. The role of causality in explainable artificial intelligence. arXiv [Internet]. 2023 Sep 18 [cited 2024 May 30]. arXiv: 2309.09901.
7. McTeer M, Applegate D, Mesenbrink P, et al. Machine learning approaches to enhance diagnosis and staging of patients with MASLD using routinely available clinical information. PLoS ONE. 2024;19(2):e0299487.
8. Nithya R, Kokilavani T, Beena TL. Balancing cerebrovascular disease data with integrated ensemble learning and SVM-smote. Netw Model Anal Health Inf Bioinf. 2024;13(1):12.
9. Kuhn M, Johnson K. Applied predictive modelling. New York: Springer; 2013.

10. Hicks SA, Strümke I, Thambawita V, et al. On evaluation metrics for medical applications of artificial intelligence. Sci Rep. 2022;12:5979.
11. Majumdar A, Bakirov R, Hodges D, et al. A multi-season machine learning approach to examine the training load and injury relationship in professional soccer. J Sports Anal. 2024;10(1):47–65.
12. Rossi A, Pappalardo L, Cintia P, et al. Effective injury forecasting in football with GPS training data and machine learning. PLoS ONE. 2018;13(7):1–15.
13. Vallance E, Sutton-Charani N, Imoussaten A, et al. Combining internal- and external-training-loads to predict non-contact injuries in football. Appl Sci. 2020;10(15):5261.

Publisher's Note