Machine learning in sports medicine: need for improvement

R Kyle Martin ,[1,2] Ayoosh Pareek ,[3] Aaron J Krych,[4,5] Hilal Maradit Kremers,[3,6] Lars Engebretsen[7,8]

The over-riding goal of a physician is to optimise the outcome for each individual patient. However, our ability to manipulate the end result at the individual level is limited by our inability to accurately predict the expected outcome of a given clinical scenario. In the age of big data, machine learning can make our predictive capability both easier and more accurate using existing registries and databases which hold the potential to dramatically change decision-making and to optimise individual outcome. The purpose of this editorial was to explore the possible uses of machine learning in sports medicine using existing knee ligament registries as an example.

[1]Department of Orthopaedic Surgery, University of Minnesota, Minneapolis, Minnesota, USA
[2]Department of Orthopaedic Surgery, CentraCare Health System, St Cloud, Minnesota, USA
[3]Department of Orthopaedic Surgery and Sports Medicine, Mayo Clinic, Rochester, Minnesota, USA
[4]Orthopedic Surgery, Mayo Clinic College of Medicine, Rochester, Minnesota, USA
[5]Sports Medicine Center, Mayo Clinic Minnesota, Rochester, Minnesota, USA
[6]Division of Epidemiology, Department of Health Sciences Research, Mayo Clinic, Rochester, Minnesota, USA
[7]Department of Orthopaedic Surgery, Oslo University Hospital, Oslo, Norway
[8]Norwegian School of Sports Sciences, Oslo Sports Trauma Research Center, Oslo, Norway

Correspondence to Dr R Kyle Martin, Department of Orthopaedic Surgery, University of Minnesota, Minneapolis, MN 56303, USA; rkylemmartin@gmail.com

Machine learning is a subset branch of artificial intelligence that uses data to make informed decisions/models without explicit programming (figure 1). Deep learning is a further subset of machine learning that uses neural networks to do the same task. Typically, once the data are acquired, significant time is spent preparing and formatting the data to be analysed, which includes removing or imputing variables which have too many missing values, standardising data for analysis and running standard statistical tests to assess relationships, such as collinearity (figure 1). Thereafter, the data are usually split into training, validation and testing data. The training data are most commonly used to select important features for the model, while the validation data are used to tune those features. Once the model is ran on blinded test data, it can be used to make predictions. These models often undergo this workflow several times to find the most precise model.

Machine learning has started to impact several medical disciplines, including orthopaedic surgery. In one example, Fontana *et al* used machine learning to determine which patients will achieve minimal clinically important difference (MCID) after total joint arthroplasty (TJA) at 2 years of follow-up.[1] Machine learning algorithms were able to predict this outcome with an area under the curve (AUC) of 0.89. The AUC of a model or test is a performance measurement which indicates how much that model/test is capable of distinguishing between classes, with values closer to 1.00 indicating better accuracy. Additionally, they identified the most important factors influencing the outcome: baseline 36-item Short Form Health Survey (SF-36) scores, unilateral revision TJA, back pain and prior knee surgery. This was made possible through a technique called ‘feature selection’, a machine learning process which narrows many variables to a small subset of the most important ones while maintaining accuracy of the model. This clinically useful registry study of 12 203 TJAs may influence decision-making by allowing accurate identification of patients who may be more amenable to non-operative management if they are less likely to reach MCID after TJA.

Using artificial intelligence to create accurate prediction tools with clinical applicability and translation holds highly impactful potential. Recently, Parvizi *et al* used a machine learning algorithm to develop a new scoring system for the diagnosis of periprosthetic hip and knee infection.[2] Whereas the previously accepted International Consensus Meeting Criteria had a sensitivity of 86.9% and a specificity of 99.5%, the new scoring system exhibited a significantly higher sensitivity of 97.7% with a similar specificity of 99.5%, and has become the new gold standard by creating an easy to use in-clinic calculator.[2] As machine learning is used more commonly by clinicians, the need for

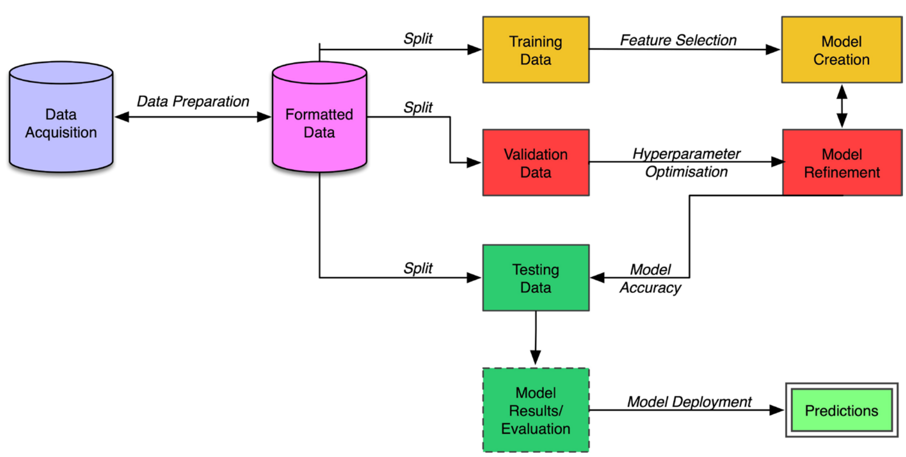

Figure 1 Typical workflow for machine learning model creation, evaluation and deployment. Data are typically split into training, validation and test sets. Training data are typically used to choose model algorithm, whereas validation sets are used for hyperparameter selection for model refinement. Model should be evaluated on test data sets, which have been blinded to model creation before deployment into a usable model for predictions.

clinically interpretable models grows. In fact, recent studies have postulated that interpretability may hinder people's ability to detect when a model has made a sizeable mistake; therefore, we must proceed with caution as clinicians involved in the creation of, and use of, these advanced models.[3]

Despite a large pool of available data, there are not yet any clinically relevant machine learning models in sports medicine. The first knee ligament registry was started in Norway in 2004 and several more now exist worldwide. These registries collect a robust volume of data, including patient demographic and injury characteristics, intraoperative findings and technical surgical details. Additionally, the collection of preoperative and postoperative patient-reported outcome measures allows capture of inferior outcome without relying on revision surgery as the sole measure. In the Scandinavian knee ligament registries alone, there are now over 70 000 patients, and the inclusion of non-operatively managed patients has also recently begun.[4] A methodology similar to the aforementioned examples can be employed to answer questions such as which patients benefit most or reach MCID after anterior cruciate ligament (ACL) reconstruction and which patients may be amenable to non-operative management of ACL tears. It is important to note that other than volume of data, other factors such as quality/completeness of data and appropriate analysis are critical for drawing appropriate conclusions.[5] Previous studies have created predictive algorithms with flawed data which can adversely affect select patient populations (such as due to racial or gender bias).[5]

In summary, machine learning can be employed (1) to predict outcomes of both surgical and non-operative sports injuries, (2) to determine which factors affect these outcomes most and (3) to allow sports medicine physicians to alter the modifiable factors which can optimise outcomes for each patient. Changing the modifiable variables would increase the probability of success and may guide treatment discussions. Though we should be cautious, this concept is already influencing patient management in other fields with the use of in-clinic models/calculators, and it is time for sports medicine to catch up. The potential to improve the care we can provide is too great to be ignored.[6]

Twitter Ayoosh Pareek @ayooshpareekmd

Contributors RKM and AP drafted and executed the manuscript, while LE contributed to the conception of of the manuscript. HMK and AJK assisted in manuscript revisions and statistical explanations. All authors contributed to the approval of the manuscript.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Not required.

To cite Martin RK, Pareek A, Krych AJ, *et al*. *J ISAKOS* 2021;**6**:1–2.

J ISAKOS 2021;**6**:1–2.
doi:10.1136/jisakos-2020-000572

ORCID iDs
R Kyle Martin http://orcid.org/0000-0001-9918-0264
Ayoosh Pareek http://orcid.org/0000-0001-8683-1697

REFERENCES

1 Fontana MA, Lyman S, Sarker GK, *et al*. Can machine learning algorithms predict which patients will achieve minimally clinically important differences from total joint arthroplasty? *Clin Orthop Relat Res* 2019;477:1267–79.
2 Parvizi J, Tan TL, Goswami K, *et al*. The 2018 definition of periprosthetic hip and knee infection: an evidence-based and validated criteria. *J Arthroplasty* 2018;33:1309–14.
3 Poursabzi-Sangdeh F, Goldstein D, Hofman J. Manipulating and measuring model Interpretability. *ArXiv* 2018;abs/1802.07810.
4 Engebretsen L, Forssblad M, Lind M. Why registries analysing cruciate ligament surgery are important. *Br J Sports Med* 2015;49:636–8.
5 Obermeyer Z, Powers B, Vogeli C, *et al*. Dissecting racial bias in an algorithm used to manage the health of populations. *Science* 2019;366:447–53.
6 Obermeyer Z, Emanuel EJ. Predicting the Future - Big Data, Machine Learning, and Clinical Medicine. *N Engl J Med* 2016;375:1216–9.