



Machine learning and conventional statistics: making sense of the differences

Christophe Ley¹ · R. Kyle Martin² · Ayoosh Pareek³ · Andreas Groll⁴ · Romain Seil⁵ · Thomas Tischer⁶

Received: 21 December 2021 / Accepted: 13 January 2022 / Published online: 2 February 2022

© The Author(s) under exclusive licence to European Society of Sports Traumatology, Knee Surgery, Arthroscopy (ESSKA) 2022

Abstract

The application of machine learning (ML) to the field of orthopaedic surgery is rapidly increasing, but many surgeons remain unfamiliar with the nuances of this novel technique. With this editorial, we address a fundamental topic—the differences between ML techniques and traditional statistics. By doing so, we aim to further familiarize the reader with the new opportunities available thanks to the ML approach.

Keywords Machine learning · Artificial intelligence · Orthopaedic surgery · Sports medicine · Statistics

Introduction

The use of artificial intelligence (AI) and machine learning (ML) with respect to orthopaedic surgery datasets has intensified over the past few years [6]. Despite the increase in studies applying these novel techniques, many orthopaedic surgeons remain unfamiliar with the concepts and how to incorporate AI into clinical practice [2]. With this editorial, we aim to clarify one commonly misunderstood aspect through exploration of the differences and similarities between classical statistical methods and AI. A fundamental understanding of how AI and ML relate to the statistical techniques traditionally employed in the orthopaedic literature can help to bridge the knowledge gap and inform

the average reader. The most important difference is that conventional statistics are model driven, while AI and ML are data driven, without an a priori understanding of the relationship between data and outcome. In AI and ML, the software recognizes patterns and creates data clusters which share common characteristics that may influence the outcome. While machine learning is technically not the same as artificial intelligence (machine learning is a subset of artificial intelligence), the two terms will be used interchangeably throughout this editorial.

Machine learning methods

The most common type of ML that is relevant for orthopaedic surgeons is called “supervised learning.” This approach consists of algorithms that analyse the relationship between “input” and “output” variables with the goal to learn how to predict a specified “output” given a set of “input” variables. The “input” variables are also commonly called “predictors” and consist of any variable in a data set that may influence or relate to an outcome. For example, in a national knee ligament registry, the “input” variables would include the patient demographic, radiographic, injury, and surgical details. In contrast, the “output” variables refer to the outcome of interest and, in the registry example, may include revision surgery, subjective outcome, or any other specified endpoints (infection, complication, length of stay, morbidity, mortality, etc.). Each patient in the registry therefore has a unique combination of “input” and “output” variables. The

✉ Christophe Ley
christophe.ley@uni.lu

¹ Department of Mathematics, University of Luxembourg, Esch-sur-Alzette, Luxembourg

² Department of Orthopedic Surgery, University of Minnesota, Minneapolis, MN, USA

³ Department of Orthopaedic Surgery, Mayo Clinic, Rochester, MN, USA

⁴ Department of Statistics, TU Dortmund University, Dortmund, Germany

⁵ Department of Orthopaedic Surgery, Centre Hospitalier Luxembourg and Luxembourg Institute of Health, Luxembourg, Luxembourg

⁶ Department of Orthopaedic and Traumatologic Surgery, Waldkrankenhaus, Erlangen, Germany

idea is that given a large enough dataset (large number of patients, each with a large number of variables), a supervised ML algorithm can identify which variable combinations are associated with the outcomes of interest.

With supervised learning, the complete dataset (including both input and output) is first divided into “training” and “test” sets. A typical approach would be to randomly assign $\approx 75\%$ of the data to the “training” set while the remaining data ($\approx 25\%$) would comprise the “test” set. Machine learning programs learn from the “training” sets and subsequently develop an algorithm to predict the “output” based on a given “input”. The accuracy of this algorithm can then be assessed using the “test” set. The data is divided to ensure proper validation of the algorithm—the “test” set should not contain data that was used to develop the algorithm in the “training” set. This approach is termed supervised learning, because the outcome of interest is identified a priori and the computer is tasked with predicting its occurrence. The ultimate goal of supervised learning is to use the algorithm to predict the outcome for new, future data.

Less common ML methods include “unsupervised learning” and “reinforcement learning”. In unsupervised learning the data is not specified as “input” or “output” variables. Instead, the AI is given all of the variables and tasked with independently finding some structure in the complete dataset. Reinforcement learning refers to a trial-and-error approach whereby the algorithm gains experience and knowledge over the course of time by constantly trying various associations. These algorithms can improve their accuracy over time in trying to achieve their goals. Reinforcement learning has for example been used to develop AI algorithms for Chess or Go game play [5], which constantly improve by playing thousands of games against themselves and are eventually unbeatable by human champion players.

Statistics versus machine learning

The recent surge of orthopaedic literature incorporating ML raises a natural question: what is the novelty compared with conventional statistical techniques such as linear or logistic regression? Indeed, traditional statistics can also ascertain a relationship between input and output and have long been used for regression and classification tasks. Further, just as with predictive ML methods, once a relationship is identified from old data, statistical approaches can subsequently be applied to new data. Some may even argue that both linear and logistic regression are themselves machine learning techniques. However, there are some important distinctions to be made between classical statistical learning and machine learning.

Statistical methods are typically top-down approaches: it is assumed that we know the model from which the data

have been generated (this is an underlying assumption of techniques like linear and logistic regression), and then the unknown parameters of this model are estimated from the data. In other words, it is assumed that we know how input variables are related to the output, which renders the interpretation of the results simple and the relationships between variables easy to understand. The potential pitfall is that the link between input and output is user chosen and may result in a suboptimal (i.e. less accurate) prediction model if the actual input–output association is not well represented by the chosen model. This may occur, for instance, if a human user chooses linear regression while in reality the relationship between input and output is non-linear, or when many input variables are involved.

Machine learning methods, in contrast, are bottom-up approaches. No particular model is assumed, but one begins with the data and an algorithm develops a model with prediction as the main goal. The resulting models are often complex, and some parameters cannot be directly estimated from the data. Instead, they are either chosen from relevant previous studies or tuned during the training in order to give the best prediction. Relative to the traditional statistical methods, ML algorithms can handle a larger number of variables, but also require a larger sample size for analysis. In other words, ML is capable of handling complex interactions in large datasets to predict outcome with greater accuracy, but the models need a greater number of input–output pairs to learn from.

A recent orthopaedic example of how ML can handle many variables with complex interactions can be found in the analysis of the Norwegian Knee Ligament Register (NKLR) [3]. In total, 24 input variables were classified as “predictors” and the outcome of interest was revision anterior cruciate ligament (ACL) reconstruction. First, the model analysed the association between the predictor variables and true outcome for $\approx 18,000$ patients. The result was an algorithm designed to predict revision surgery. The performance of the algorithm was tested on the remaining ≈ 6000 patients in the NKLR. Further, through a technique known as feature selection, the large number of variables initially included in the model (24) were pared down to the minimum number necessary for prediction without sacrificing accuracy. This resulted in an algorithm capable of revision prediction that only requires the input of *five* variables. The ability to realize the complex interactions between all the variables while also eliminating those with minimal contribution to outcome prediction is a hallmark of ML techniques.

The primary distinguishing feature of ML methods is the fact that they are data driven rather than user chosen with the goal of accurate prediction. This prevents the error of applying the wrong statistical model to the dataset which may limit accuracy. These models are not without limitations however, especially regarding clinical utility. Since the

focus of ML is primarily on prediction accuracy rather than on identifying relationships, the biggest downside to ML approaches relates to interpretability of the models, which explains why some models are termed black-box models (e.g. neural networks). In striving for optimal prediction accuracy, an understanding of how the algorithm determined the prediction may be sacrificed for the black-box models.

Predicting injury risk: an example of two techniques

Both traditional statistical techniques and ML can be used to predict the occurrence of an event. For the purposes of illustration, we will walk through an example of both approaches to the prediction of knee injury risk in a soccer player. Two risk factors (training load and history of previous injury) will be used for analysis. Estimating injury risk is a binary classification task (YES or NO), and in the following paragraphs it is described how logistic regression (traditional statistics) and random forest classification (ML technique) can tackle this problem.

In logistic regression, the model is chosen by the user. In this case, a model equation is created where the probability of sustaining an injury is linked to the input via a mathematical function. The baseline risk corresponding to zero training load and zero previous injuries is defined by a specific parameter. If there is some interaction between training load and previous injuries (for example, reduced training load after an injury), this can be added to the model as well. Several parameters are then estimated from the available data and indicate how much each predictor contributes to the overall injury risk. This method is relatively straightforward in the case of only two risk factors to consider. However, a more realistic scenario is one with a larger number of possible contributing predictors such as age, playing position, playing surface, shoe type, body weight, height, weather conditions, results of physical testing, morphological parameters, and many more. In that case, the situation may quickly become extremely complex. All possible pairs of predictors and their potential interactions (and maybe even non-linear effect types) must be considered, making it difficult to detect and quantify individual contributions given the magnitude of the equation. The advantage of logistic regression lies in the fact that, once the model is defined, the process of calculating injury risk for each new individual is straightforward, easy to understand, and reproducible.

Machine learning can address the matter of complexity in this scenario. In this example, a method called *random forest*

[1] approach can be used to estimate injury risk. As the name suggests, a random forest consists of several individual classification trees like the one depicted in Fig. 1.

To estimate the risk for a given soccer player for this single classification tree, one starts at the top of the tree with the first risk factor, “Previous injuries.” Working down through the algorithm, each split either leads to the next decision point (split) or to a node (also called leaf) denoting the estimated risk score along the bottom of the figure. The probability of sustaining an injury is represented by the black shaded portion of the leaf and is highest on the far right of the figure.

The injury risk for the entire random forest is obtained as a combination of the results from the individual trees. These individual trees are visually easy to understand and automatically take interactions into account due to the cascading nature. No user-based model choice needs to be done beforehand as all interaction terms are data driven. This allows the individual trees and resulting random forest to effectively manage a large number of predictors. Although the interpretability of an entire forest is very difficult relative to the individual trees, the ability to predict injury risk is greatly improved owing to the high accuracy of the model based on the complex interactions between variables.

Conclusion

The biggest difference between traditional statistics and AI/ML is the approach to the model generation. In statistics, a mathematical model is created by the user while in ML, the model is essentially created by the algorithm based on the available data. The result is that ML is in general superior when handling many variables—especially if there are complex interactions between these variables. While better suited for handling complex datasets, ML approaches often sacrifice interpretability relative to standard statistics since the goal is to optimize prediction accuracy. Interpretable ML or explainable AI [4] represent recent research approaches aimed at providing solutions for this weak spot, meaning that interpretability may improve with future models. The impact of ML on the orthopaedic literature will continue to increase and it is important for clinicians to understand the applications, limitations, and interpretation. Otherwise, clinical translation of new knowledge may be inhibited, slowing the growth of the speciality. The more the orthopaedic community can embrace this novel approach, the sooner its potential will be unleashed.

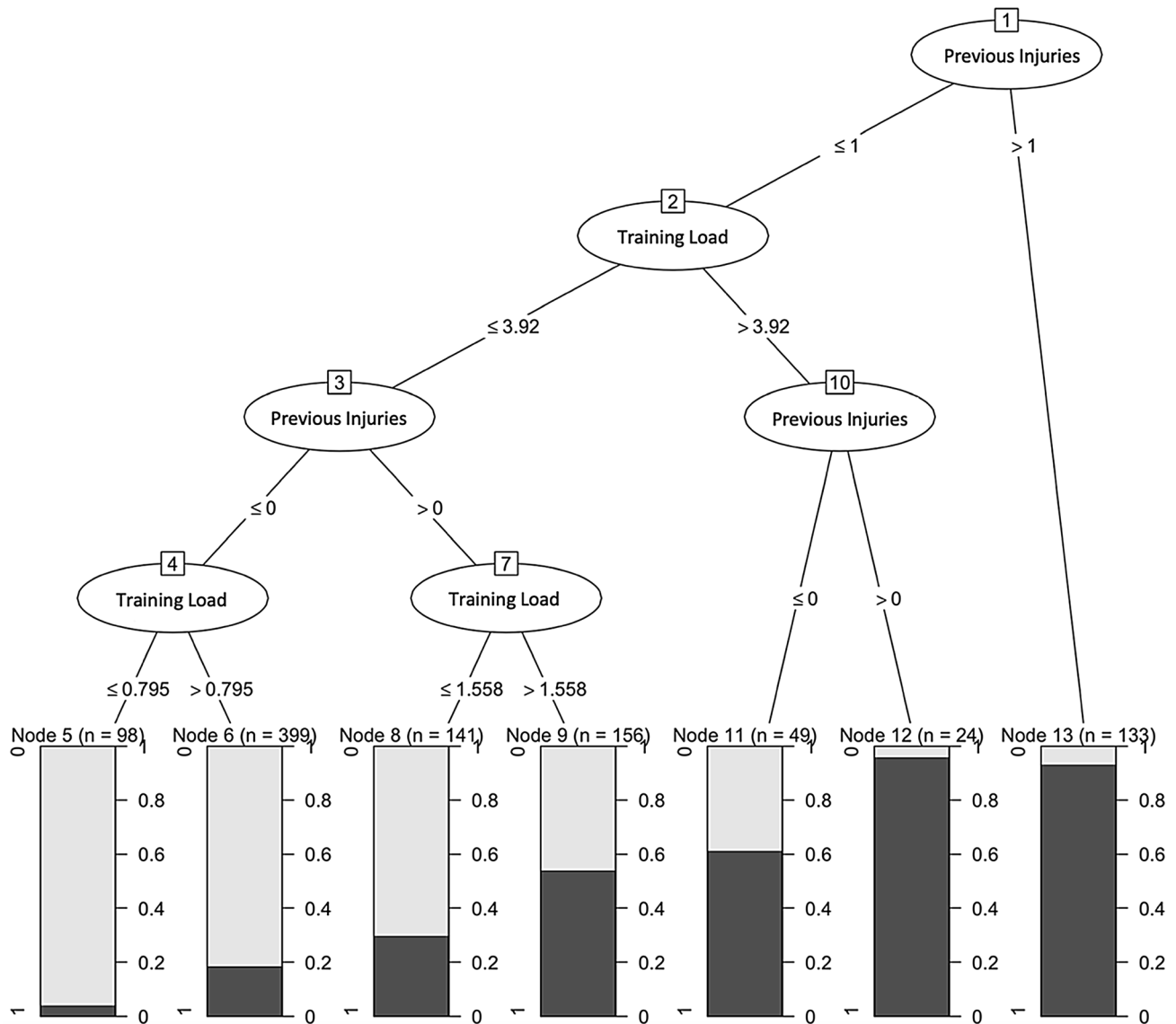


Fig. 1 Example of a single classification tree from the random forest algorithm built for estimating the risk of injury for soccer players based on the predictors Training load and Previous injuries. The individual branches display the threshold values for each split, that is, the

values, associated with a predictor, above and below which the risk is different; these values are chosen by the algorithm on the basis of the available training data. Node 5 shows minimal injury risk (black shaded portion), while nodes 12 and 13 show very high injury risk

Acknowledgements Not applicable.

Author contributions All listed authors have contributed substantially to this work: CL, RKM and TT performed the primary manuscript preparation. AP, AG, and RS assisted with editing and final manuscript preparation. All authors read and approved the final manuscript.

Funding No funding was needed.

Declarations

Conflict of interest The authors declare that they have no conflict of interests.

References

1. Breiman L (2001) Random forests. *Mach Learn* 45:5–32
2. Martin RK, Ley C, Pareek A, Groll A, Tischer T, Seil R (2021) Artificial intelligence and machine learning: an introduction for orthopaedic surgeons. *Knee Surg Sports Traumatol Arthrosc*. <https://doi.org/10.1007/s00167-021-06741-2>
3. Martin RK, Wastvedt S, Pareek A, Persson A, Visnes H, Fenstad AM, Moatshe G, Wolfson J, Engebretsen L (2021) Predicting anterior cruciate ligament reconstruction revision: a machine learning analysis utilizing the Norwegian Knee Ligament Register. *J Bone Joint Surg Am*. <https://doi.org/10.2106/JBJS.21.00113>

4. Molnar C (2019) Interpretable machine learning: a guide for making black box models explainable. <https://christophm.github.io/interpretable-ml-book/>
5. Silver D, Schrittwieser J, Simonyan K, Antonoglu I, Huang A, Guez A, Hubert T, Baker L, Lai M, Bolton A, Chen Y, Lillicrap T, Hui F, Sifre L, van den Driessche G, Graepel T, Hassabis D (2017) Mastering the game of Go without human knowledge. *Nature* 550:354–359
6. Van Eetvelde H, Mendonça LD, Ley C, Seil R, Tischer T (2021) Machine learning methods in sport injury prediction and prevention: a systematic review. *J Exp Orthop* 8:27. <https://doi.org/10.1186/s40634-021-00346-x>

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