Injury Prediction in Competitive Runners with Machine Learning

Lovdal, Sofie; den Hartigh, Ruud; Azzopardi, George

[Link to publication in University of Groningen/UMCG research database](#)

Human Kinetics

Injury Prediction in Competitive Runners With Machine Learning

S. Sofie Lövdal, Ruud J.R. Den Hartigh, and George Azzopardi

Purpose: Staying injury free is a major factor for success in sports. Although injuries are difficult to forecast, novel technologies and data-science applications could provide important insights. Our purpose was to use machine learning for the prediction of injuries in runners, based on detailed training logs. ***Methods:*** Prediction of injuries was evaluated on a new data set of 74 high-level middle- and long-distance runners, over a period of 7 years. Two analytic approaches were applied. First, the training load from the previous 7 days was expressed as a time series, with each day's training being described by 10 features. These features were a combination of objective data from a global positioning system watch (eg, duration, distance), together with subjective data about the exertion and success of the training. Second, a training week was summarized by 22 aggregate features, and a time window of 3 weeks before the injury was considered. ***Results:*** A predictive system based on bagged XGBoost machine-learning models resulted in receiver operating characteristic curves with average areas under the curves of 0.724 and 0.678 for the day and week approaches, respectively. The results of the day approach especially reflect a reasonably high probability that our system makes correct injury predictions. ***Conclusions:*** Our machine-learning-based approach predicts a sizable portion of the injuries, in particular when the model is based on training-load data in the days preceding an injury. Overall, these results demonstrate the possible merits of using machine learning to predict injuries and tailor training programs for athletes.

Keywords: data science, distance running, training load, predictive modeling, XGBoost

Staying healthy and injury free is one of the most important factors for optimal performance in sports.[1,2] Therefore, for decades, researchers and practitioners across different sports have collected data on training loads of athletes and the occurrence of injuries.[3–7] Recent years have witnessed a growth in technologies and machine learning applications, which can be employed to make predictions about future performance, injuries, and thereby improve data-driven guidance in sports.[8–10]

In the current study, we use a supervised machine learning approach, which relies principally on presenting numerous examples of data points from each group of interest. In the case of predicting injuries, this means feeding the learning algorithm examples of, for instance, training weeks that lead to injury as well as examples of training weeks that do not lead to injury. Contrary to traditional techniques such as linear regression, supervised machine learning can model complex patterns between many variables.[11] Furthermore, it does not make any assumptions about the type or degree of nonlinearity between the independent and dependent variables, or the underlying distributions of independent variables. The resulting predictive model, therefore, fits itself to the available data, and does not use a predefined model to evaluate how well the data fits to it.

So far, the majority of research applied simple or advanced statistical (regression) methods to predict injuries, with mixed results.[4,5,12] For instance, Raya-Gonzalez et al[12] conducted generalized estimating equation analysis to examine the associations between different load-markers and noncontact injuries in the subsequent week for soccer players. They achieved an area under the curve (AUC[I]) below 0.50, based on which they concluded that

internal load markers have poor predictive capacity of injuries. Interestingly, the few machine learning attempts in the past years have been more successful in predicting injuries. Most of these studies were conducted in team sports in which data on workloads and injuries are collected on a daily basis.[13–15] For example, Rossi et al[13] created an injury predictor for soccer players. Specifically, using player global positioning system data to extract details from training, the researchers considered both current and previous external training load to forecast whether or not a player would get injured. Their decision tree-based classifier could predict 80% of the injuries with a precision of 50%, resulting in an AUC of 0.76. This result is in line with recent research on Australian rules football players, which found AUC scores varying between 0.75 and 0.80 across a season.[15]

While promising first steps are made on the application of machine learning to predict injuries several challenges remain to be acknowledged and addressed. First, acute injuries are often unpredictable by definition. Such injuries occur relatively frequently in team sports like soccer, meaning that a portion of injuries will always remain unpredictable. Related to this, individual endurance sports such as running, in which overuse injuries are the most prevalent, may be most suited for accurate injury prediction.[16] Furthermore, from an analytic perspective, methods applied so far expressed workload as some form of aggregation, such as the duration of training expressed over a longer time window. Although such aggregations may well capture the accumulated external load put on the body of an athlete, the very important sequence property of a time series that the training of an athlete represents is lost. In other words, the sequence and time lag of the exercises done in the days or weeks prior to an injury are not taken into account, which may be crucial for explaining the occurrence of an injury.[17,18] Finally, to the best of our knowledge, studies so far have used relatively small data sets (covering between one season and 2 y, often including relatively few injury events). For instance, the data set by Rossi et al[13] included only 23 injuries. In general,

Lövdal and Azzopardi are with the Bernoulli Inst for Mathematics, Computer Science and Artificial Intelligence, Faculty of Science and Engineering, and Den Hartigh, the Dept of Psychology, Faculty of Behavioral and Social Sciences, University of Groningen, Groningen, the Netherlands. Azzopardi (g.azzopardi@rug.nl) is corresponding author.

fewer events go at the expense of the reliability and generalizability of the machine learning model.

Our study addresses the challenges above, and aims to model the prediction of injuries in runners with data collected across several years, thereby including many injury events. To test the relevance of predictions on lower levels of aggregation, we developed a model in which the focus lies on the training load data in the days before the injury (microlevel), and another one in which the focus lies on load data in the weeks before the injury (macrolevel). We tested these models to answer the research question: How accurately can we predict whether the next training session will result in an injury? Furthermore, we provide insights on the factors that contribute the most in each of the 2 predictive models.

Methods

Subjects

The data set consists of a detailed training log from a Dutch high-level running team over a period of 7 years (2012–2019). We included the middle- and long-distance runners of the team, that is, those competing on distances between the 800 m and the marathon. This design decision is motivated by the fact that these groups have strong endurance-based components in their training, making their training regimes comparable. The head coach of the team did not change during the years of data collection. The data

set contains samples from 77 runners, of whom 27 are women and 50 are men. At the moment of data collection, they had been in the team for an average of 3.7 years. Most athletes competed on a national level, and some also on an international level. The study was conducted according to the requirements of the Declaration of Helsinki and was approved by the ethics committee of the second author's institution (research code: PSY-1920-S-0007).

Design

The training log contains detailed information about each training session filled in by each athlete. Running training is expressed in terms of the number of kilometers covered in different intensity zones, and alternative training (cross-training) is logged together with the type (such as cycling or swimming) and duration. This data are collected by global positioning system watches with heart rate monitors that log the training duration, distance covered, and heart rate. The last type of training logged is strength training, and every session also contains subjective information about how well the athlete felt before the start of the session (perceived recovery), how exhausted the athlete felt upon completion of the training (perceived exertion), and how well the athlete thought the training went (perceived training success) (Table 1).

Injuries can be extracted from the so-called "flags" in the training log. These flags are defined as "unable to complete the

Table 1   Features Used for the Day (Left) and Week (Right) Approaches, Together With Their Typical Value Range

No	Day feature	Range	No	Week feature	Range
1	Number of sessions	[0, 2]	1	Number of sessions	[0, 14]
2	Total distance	[0.0, 25.0]	2	Number of rest days	[0, 7]
3	Sum of distance in Z3–Z4	[0.0, 15.0]	3	Total distance	[0.0, 150.0]
4	Sum of distance in Z5, T1, and T2	[0.0, 10.0]	4	Max distance (1 day)	[0.0, 30.0]
5	Distance sprinting	[0.0, 1.5]	5	Total distance Z3–5, T1–T2	[0.0, 30.0]
6	Number of strength sessions	[0, 1]	6	Sessions in Z5–T1–T2	[0, 3]
7	Hours alternative training	[0.0, 3.0]	7	Sessions in Z3 or faster	[0, 4]
8	Perceived exertion	[0.0, 1.0]	8	Total distance Z3–Z4	[0.0, 20.0]
9	Perceived training success	[0.0, 1.0]	9	Max distance Z3–Z4 (1 day)	[0.0, 15.0]
10	Perceived recovery	[0.0, 1.0]	10	Total distance Z5–T1–T2	[0.0, 20.0]
			11	Max distance Z5–T1–T2	[0.0, 10.0]
			12	Hours alternative training	[0.0, 20.0]
			13	Number of strength training sessions	[0, 7]
			14	Average exertion	[0.0, 1.0]
			15	Minimum exertion	[0.0, 1.0]
			16	Maximum exertion	[0.0, 1.0]
			17	Average training success	[0.0, 1.0]
			18	Minimum training success	[0.0, 1.0]
			19	Maximum training success	[0.0, 1.0]
			20	Average recovery	[0.0, 1.0]
			21	Minimum recovery	[0.0, 1.0]
			22	Maximum recovery	[0.0, 1.0]
			67	Total distance week 1/week 2	[0.0, 3.0]
			68	Total distance week 0/week 1	[0.0, 3.0]
			69	Total distance week 0/week 2	[0.0, 3.0]

Note: Z1–Z5 represent different heart-rate zones where Z1 is easy aerobic effort and Z5 is close to maximum heart rate. T1 and T2 are long and short track intervals, which are typically done at high intensity. The day approach uses a set of 10 features per day, and the week approach uses a set of 22 aggregated features per week along with the bottom 3 features that describe relative increase in volume.

scheduled session due to injury." They cover both cases where an athlete starts a session but interrupts it due to injury, and cases where the athlete skips the session due to injury.

In our data set, injuries are reflected by all records flagged with injury from the training log, whereby it was required that the athletes were training injury free 3 weeks prior to the session flagged with injury. For the healthy events, we demanded that the athlete is fully fit 3 weeks before and 3 weeks after the event day. Moreover, events that contained missing or anomalous data were removed from our data set.

Finally, injury events shortly following (within 3 wk of) a new injury have been filtered out, as they are considered to correspond to the same injury. This leaves us with a total of 74 athletes (27 women and 47 men) who have 42,183 healthy and 583 injury events for the microlevel model and 42,223 healthy and 575 injury events for the macrolevel model (see next section for the details about these models, which we refer to as the day approach and week approach, respectively). In both cases, the fraction of injury events is approximately 1.4%. The number of injuries per athlete varies between 0 and 35.

Feature Construction

The machine learning algorithm that we use takes as input a set of feature vectors (data describing a training setup) and the corresponding label (injury or healthy). In an iterative approach, the algorithm determines a predictive model that best maps the input feature vectors to the corresponding labels.

We investigate 2 different approaches to capture the training load leading up to an injury. The day approach considers 1 week before the event (injury or healthy) and considers each day individually. The week approach considers 3 weeks before the event and aggregates each training week by a set of features. For both approaches, we normalize the independent variables using the z score transform for each athlete individually, based on the mean and SD of the healthy events of that athlete.

In the day approach, a feature vector is constructed by expressing the week before the injury or healthy event as a series of days described by ten features per day (ie, 70 features in total, Table 1). We count the days starting from 0, so the day before the event is seen as day 0, 2 days before the event as day 1 and 7 days before the event is day 6 (Figure 1A).

The week approach considers 3 weeks before the injury or healthy event, where we summarize the training load on week level. Each week leading up to an event is described by 22 features, listed in Table 1. Furthermore, as illustrated in Figure 1B, 3 features were added describing change in weekly mileage (total distance covered by running per week[19]), leading to 69 features per data point.

Data-Driven Model

The machine learning algorithm chosen for this research is called Extreme Gradient Boosting, or XGBoost,[20] having provided the algorithm behind many top-ranking machine learning classifiers over the past few years.[21–23] For more technical details on XGBoost, we refer to the work of Chen and Guestrin.[20]

As our data set is highly unbalanced (ie, many more healthy events than injury events), and a machine learning classifier typically needs to be trained on a balanced data set in order to avoid bias toward the majority class, we implement a bagging approach. Here, multiple models are trained on balanced subsets of the training data. For a given test event (injury or healthy), the prediction is then calculated as the mean of all predictions by the participating models. In this way, we can get a larger representation of healthy examples included than when training a single model. We do not fine tune the participating models. In practice, we use a standard small value for the learning rate, we choose between 2 random values for 2 of the hyperparameters (Table 2), and use default values for the remaining ones.[24]

Each balanced subset is determined by randomly selecting, with replacement, the same number of injury and healthy events from all athletes in the training set. This is to avoid having the events of a few high-risk athletes dominating the others. More specifically, each model is trained and calibrated using a sample size of 2048 injury and 2048 healthy events. We train 9 models
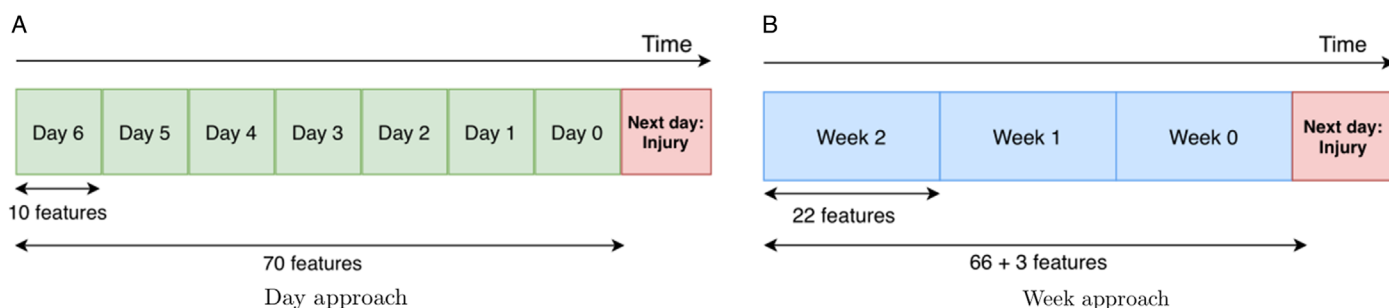

Figure 1 — Structure of the feature vectors. Day approach: A data point is described by the training-defining features over the 7 days before the day of prediction. Given that each day is described by 10 elements, in the day approach, a data point is defined as a 70-element feature vector. Week approach: Each week is described as a summary of 22 features. The final feature vector is a sequence of 3 sets of such features over 3 weeks along with 3 features describing the change in total distance covered from week to week.

Table 2   Key Parameters Used When Training the XGBoost Model[24]

Parameter	Value	Function
learning_rate	0.01	Step size for correcting the model
max_depth	{2, 3}	Maximum depth of a tree
n_estimators	{256, 512}	Total number of trees that are built for the model

using the data of the first 64 athletes that joined the team. Calibration is a technique that transforms the output of a machine learning model, such that the transformed output matches the observed distribution of a class in the training set. This means that the calibrated output will be able to indicate the fraction of injuries, which have a lower (or higher) score than what is outputted by the model. For example, if the calibrated model outputs a score of 0.80, we know that 80% of the injuries have a lower score than the event under consideration, and therefore the given event has a very high risk of being an injury.[25] In particular, we use Platt scaling to calibrate all involved XGBoost models.
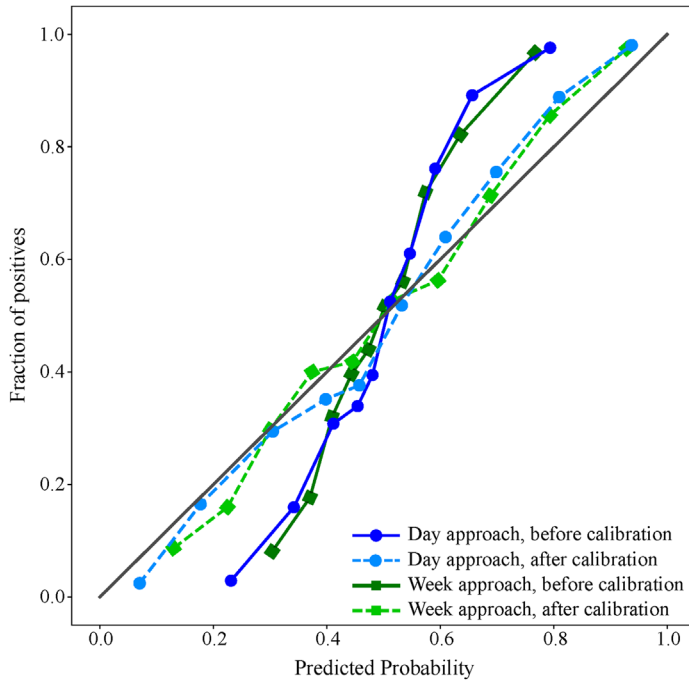

Figure 2 — Averaged reliability curves, across 5 experiments, before and after calibrating the bagged XGBoost models of the day and week approaches. The black solid diagonal line indicates a perfectly calibrated model and is used as a reference to compare the other curves.
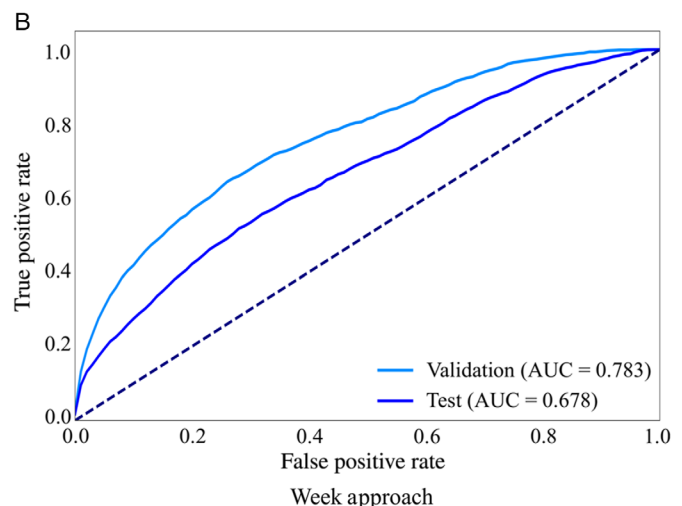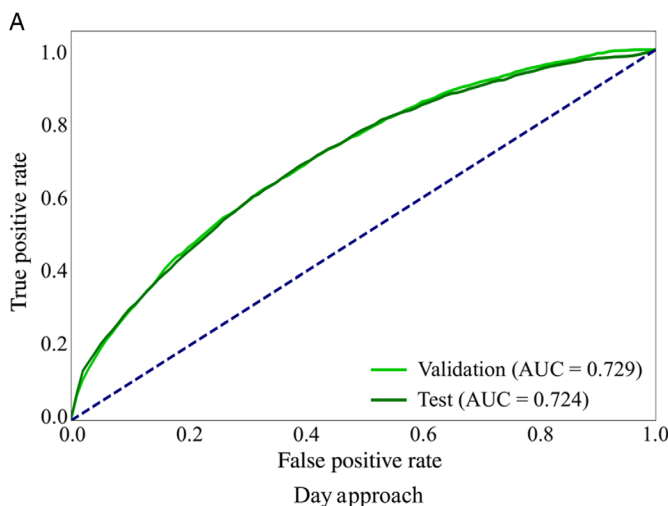
We use temporal evaluation by holding out the data of the 10 athletes that most recently joined the team as a test set. We then test the model by applying the predictive model on all events of the test athletes. In order to test the generalizability of our approach, we also create a validation set by randomly selecting (without replacement) a subset from the training data with the same distribution as the test set, namely 50 injury and 2994 healthy events.

A prediction using our calibrated bagging approach is reported as a real value in the range [0, 1]. We set the threshold for injury as the one that leads to the minimum difference between the specificity and sensitivity on the validation set. Then, the test events that achieve a score equal to or above the threshold are labeled as injury, otherwise they are considered healthy. A receiving operator characteristic (ROC) curve describes the fraction of true positives (injuries detected) versus the corresponding rate of false positives, as the decision threshold is varied. Moreover, the AUC of the ROC describes the performance of the model by a single value. The higher the AUC the higher is the degree of separability between instances that lead to injury or not. An AUC of 1.0 would describe a perfect prediction model, and 0.5 corresponds to random guessing. Hence, the closer the AUC is to 1 the better the prediction model, and AUC scores of 0.7 and above are considered as having strong effects in the field of sports sciences.[15,26]

Results

In this section, we present the average results across 5 experiments that we conducted.[II] Figure 2 shows the average reliability curves of the bagging models before and after calibrating the involved XGBoost classifiers of the day and week approaches. The roughly straight diagonal lines obtained by the bagging models with calibrated classifiers indicate the reliability of their results, in that they neither underestimate nor overestimate the risks. In Figure 3, we present the validation and test ROC curves of the calibrated day and week approaches. These are generated based on the mean score determined by the 9 participating XGBoost models, with the pairs of true and false positive rates plotted along the curves. We obtain the average AUC scores of 0.729 and 0.724 for the validation and test sets of the day approach, and AUCs of 0.783 and 0.678 for the validation and test sets of the week approach, respectively. The

Figure 3 — Averaged receiver operating characteristic curves for the validation and test sets for (A) the day approach and (B) the week approach. The dashed line corresponds to random guessing between injury and noninjury (AUC = 0.5). AUC indicates area under the curve.

consistent AUCs obtained for the validation and test sets demonstrate the ability of our approach to generalize what is learned from the training athletes to the events of new (unseen) test athletes within the same running team.

To provide a closer look at the performance of our predictive model, we report additional results in Table 3. The day approach outperforms the week approach on all measures. Sensitivity (recall) or true positive rate describes the fraction of injury samples that is detected in the test set. For each approach, we report results for a particular threshold, which we determine as the mean of all thresholds that yield the minimum differences between the sensitivity and specificity across the 5 experiments on the validation data. For such thresholds, which are 0.448 and 0.476 for the day and week approach, respectively, our model achieves an average sensitivity of 58.4% for the day approach and 50.4% for the week approach. The respective average specificities, or true negative rates, are 74.1% and 74.6%.

The XGBoost algorithm also determines various measurements about the relevance or importance of each input feature with respect to the output label. Figure 4 shows the 20 most important features based on the feature importance indicator "total gain" for the day and week approach, respectively. They are averaged across the 9 participating models and across the 5 experiments. "Total gain" is a quantity that describes how often a feature is considered in the model together with how distinctive it is in separating noninjury from injury data points.[24] In other words, variables with high feature importance are queried often in the model, and the queries made on these features result in a relatively clear separation between injury and noninjury data points.

Figures 5A and 5B display the correlations between the individual features (after z score normalization) and the dependent variable (injury or healthy). The correlations can be considered low for both the day and week approaches, as they vary between $r = -.01$ and $r = .07$ or $r = -.04$ and $r = .07$, respectively. It should be noted that the correlations are not necessarily reflective of total gain reported in the feature importance plots, which becomes visible when comparing the correlation results (Figures 5A and 5B) with the results on the feature importance (Figure 4).

Discussion

The current study is among the first to propose data-driven predictive models of injuries in running. The models are configured with a well-founded machine learning algorithm: XGBoost.[20,21] The results, in particular those of the day approach, demonstrate the effectiveness of the proposed approach to predict injuries.

The finding that the day approach performs best, suggests that our model with load indicators from the previous week has the highest predictive capacity of an injury. Hence, the sequence and time lag of the exercises done in the days before the injury seem to provide important input, which is in line with previous literature demonstrating the relevance of acute training load in predicting sports injuries.[27] This suggests that the monitoring and analysis of daily training loads are important to include in the machine learning pipeline, and need to be taken into account when modeling injury occurrences and establishing (weekly) training programs.[18]

In order to gain insight into the features that mostly affect the prediction of injuries, we analyzed the "total gain" of each feature, which is determined by the learning process of the XGBoost models. This is a measurement based on how often the XGBoost models use the concerned feature to come to a decision, and it is determined by taking the interactions between all involved features into account. For comparison purposes, we also determined the Pearson correlation between each individual feature and the injury label. It can be noted that the individual correlations are weak (Figures 5A and 5B), which suggests that the build up to an injury cannot be reduced to one or a few independent variables. Instead, it is likely due to a complex interaction among various variables as determined by the XGBoost model.[28,29] For instance, as can be observed in Figure 4, the feature "km Z5-T1-T2.6" is the most important feature in our model of the day approach. This feature would be overlooked when using a classical technique, because the correlation between "km Z5-T1-T2.6," and the injury label is almost neutral ($r = .01$; Figure 5A). This demonstrates the ability of our model to discover complex relationships between features.

Taken together, our machine learning-based system generated relatively good predictions of injuries in running athletes. A major strength of our study is that we used a rich data set that covers a period of 7 years, including almost 600 injuries. Such a high number of injury events allowed for learning a robust model, which showed high generalizability to unseen athletes within the same running team, as manifested by very similar validation and test AUCs. This is in contrast to previous studies in sports that often trained the models on smaller data sets.[10,13,15]

For future research, possibilities to further improve the performance of the model may be explored. For instance, other supposedly relevant factors such as footwear and biomechanics of the athletes, could be included in the data set and thereby the resulting model.

Practical Applications

The proposed system can be considered as a computer-aided tool that may be used to assist coaches in regulating the training load for athletes. More specifically, the system is capable of determining a risk score between 0 and 1, which the coach can then use together with previous injury history of an athlete to assess the effect of the training load. In doing so, a coach could take extra caution with the training setup of an athlete when the system gives a high injury risk score. For instance, the coach could prescribe extra rest to let the athlete in question recover, or assign resources to further investigate their (pre)injury status.

Table 3 Mean (SD) Test Scores Obtained by 5 Experiments, With a Threshold of 0.448 for the Day Approach and 0.476 for the Week Approach

Approach	Specificity	Sensitivity	AUC
Day	0.741 (0.02)	0.584 (0.05)	0.724 (0.01)
Week	0.746 (0.04)	0.504 (0.05)	0.678 (0.01)

Note: These thresholds are the average values that correspond to the minimum difference between the specificity and sensitivity across the 5 experiments on the validation data. Abbreviation: AUC, area under the curve.
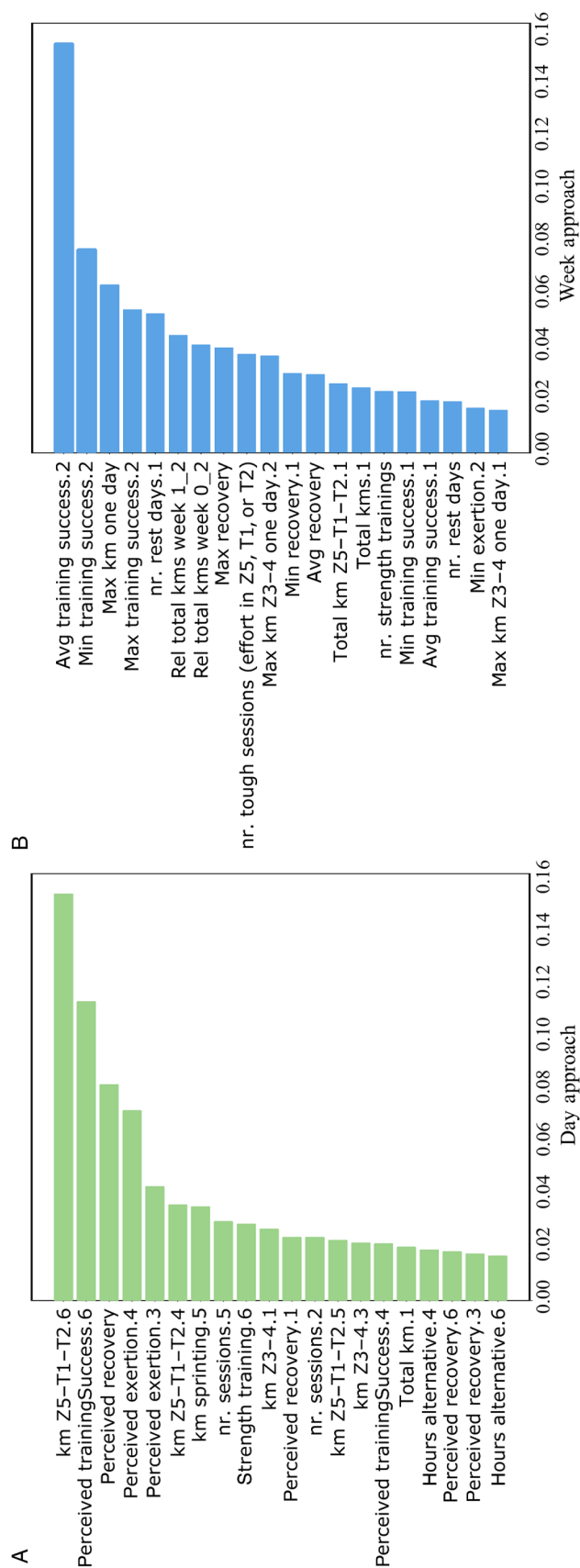
Figure 4 — The top 20 most important features according to the feature importance criterion "total gain," expressed as their relative contribution to the XGBoost model. Please note that here the features describing each day or week start counting from 0, so "perceived recovery" in (A) describes how well the athlete felt before the training the day before the injury, and "perceived recovery.1" describes the same feature 2 days before the injury.

Figure 5A — Pearson correlation between every feature in the day approach and the injury label. The number after each feature name indicates which day it represents, with no number being the day before the injury, number 1 being 2 days before the injury, and so on.

Figure 5B — Pearson correlation between every feature in the week approach and the injury label. The number after each feature name indicates which day it represents, with no number being the day before the injury, number 1 being 2 days before the injury, and so on.
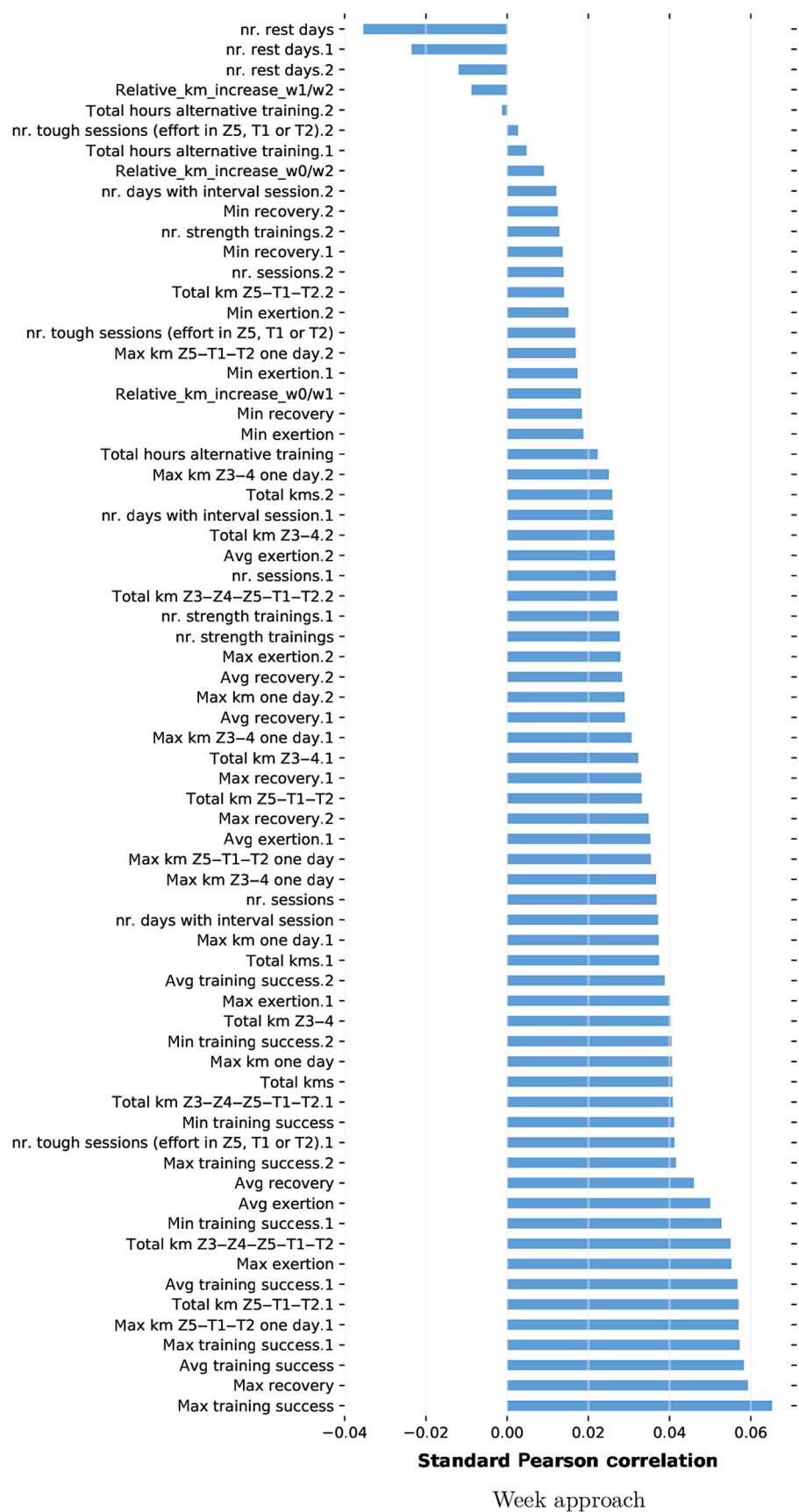
When using the system, the number of correctly detected injuries (ie, sensitivity) can be increased at the cost of obtaining more false positives, hence reducing the specificity. This can be accomplished by varying the decision threshold that can be chosen based on the user requirements. Looking at the ROC curve of the day approach (Figure 3), it can be noted that, theoretically, around 30% of the injuries can be detected with a false positive rate of only 10%. This means that the system could indicate when the training load should be considered with caution, since obtaining an injury generally has a much longer recovery time than what you lose from taking a rest day. However, even when the athlete takes a rest day in case of a high injury risk score, it does not mean that the athlete is immediately in the safe zone, especially considering the sneaky build-up of overuse injuries.[17,18] Choosing a threshold is, therefore, up to the athletes and coaches to do. Selecting a low threshold could, for example, be desirable close to a main competition event of the season.

The difference in practical applications between the day and the week approach is that the latter represents accumulated (chronic) strain on the body over a longer period (3 wk), with features being a summary of the workload. The day approach, on the other hand, considers more detail but on a shorter period (1 wk) with each day being represented individually. Note that, as shown in Figure 4, the most important features of the day approach primarily relate to activities 4 to 7 days before the injury, whereas the most important features for the week approach primarily relate to 2 or 3 weeks before the injury. Hence, signs of an injury may be extracted days-to-weeks before it occurs. In this regard, future work may investigate further early signs for the prediction of injury with more than one training day in advance.

Finally, our proposed pipeline can be adapted by other running teams that collect comparable data. More generally, we hope that our work can serve as an example of how data-driven modeling could contribute to sports science. This discipline is benefiting from the rapid development of wearables, which enable the collection of abundance of data with different types of sensors and with high frequency.[8,9,31]

Conclusions

Our predictive model based on the XGBoost algorithm and bagging is effective in predicting injuries based on training load data of middle- and long-distance runners. The day approach outperforms the week approach, which suggests that injuries can largely be predicted based on our model with load indicators in the days before the injury. As a byproduct, the proposed approach indicates the relevance of all independent variables in the decision-making process.

To conclude, the data-driven approach that we propose is general, which means that it can also be adapted to other sports. This may contribute to novel and improved ways of data-driven guidance in the sports field.

Notes

I. AUC indicates the tradeoff between the true and false positive rates as one changes the decision criteria. The closer the AUC is to 1 the better the system is. A value of .50 corresponds to no discriminating ability.[30]

II. The data set used in this study and the Python code to replicate our results are available through the following link: https://doi.org/10.34894/UWU9PV

References

1. Raysmith BP, Drew MK. Performance success or failure is influenced by weeks lost to injury and illness in elite Australian track and field athletes: a 5-year prospective study. *J Sci Med Sport*. 2016;19(10):778–783. PubMed ID: 26839047 doi:10.1016/j.jsams.2015.12.515

2. D'souza D. Track and field athletics injuries–a one-year survey. *Br J Sports Med*. 1994;28:197–202. PubMed ID: 8000821 doi:10.1136/bjsm.28.3.197

3. Soligard T, Schwellnus M, Alonso J-M, et al. How much is too much? (Part 1) International Olympic Committee consensus statement on load in sport and risk of injury. *Br J Sports Med*. 2016;50(17):1030–1041. PubMed ID: 27535989 doi:10.1136/bjsports-2016-096581

4. Brink MS, Visscher C, Arends S, Zwerver J, Post WJ, Lemmink KAPM. Monitoring stress and recovery: new insights for the prevention of injuries and illnesses in elite youth soccer players. *Br J Sports Med*. 2010;44(11):809–815. PubMed ID: 20511621 doi:10.1136/bjsm.2009.069476

5. Cross MJ, Williams S, Trewartha G, Kemp SPT, Stokes KA. The influence of in-season training loads on injury risk in professional rugby union. *Int J Sports Physiol Perform*. 2016;11(3):350–355. doi:10.1123/ijspp.2015-0187

6. Dennis R, Farhart R, Goumas C, Orchard J. Bowling workload and the risk of injury in elite cricket fast bowlers. *J Sci Med Sport*. 2003;6(3):359–367. PubMed ID: 14609154 doi:10.1016/S1440-2440(03)80031-2

7. Huxley DJ, O'Connor D, Healey PA. An examination of the training profiles and injuries in elite youth track and field athletes. *Eur J Sport Sci*. 2014;14(2):185–192. PubMed ID: 23777449 doi:10.1080/17461391.2013.809153

8. Jaspers A, Op De Beéck T, Brink MS, et al. Relationships between the external and internal training load in professional soccer: what can we learn from machine learning? *Int J Sports Physiol Perform*. 2018;13(5):625–630. doi:10.1123/ijspp.2017-0299

9. Knobbe A, Orie J, Hofman N, Burgh B, Cachucho R. Sports analytics for professional speed skating. *Data Min Knowl Disc*. 2017;31(6):1872–1902. doi:10.1007/s10618-017-0512-3

10. Naglah A, Khalifa F, Mahmoud A, et al. Athlete-customized injury prediction using training load statistical records and machine learning. *2018 IEEE Int Symp Signal Process Inf Technol*. 2018:459–464. doi:10.1109/isspit.2018.8642739

11. Alpaydin E. *Introduction to Machine Learning*. Cambridge, MA: MIT Press; 2009.

12. Raya-Gonzalez J, Nakamura FY, Castillo D, Yanci J, Fanchini M. Determining the relationship between internal load markers and noncontact injuries in young elite soccer players. *Int J Sports Physiol Perform*. 2019;14(4):421–425. doi:10.1123/ijspp.2018-0466

13. Rossi A, Pappalardo L, Cintia P, Iaia FM, Fernàndez J, Medina D. Effective injury forecasting in soccer with GPS training data and machine learning. *PLoS One*. 2018;13(7):e0201264. PubMed ID: 30044858 doi:10.1371/journal.pone.0201264

14. Talukder H, Vincent T, Foster G, et al. Preventing in-game injuries for NBA players. Paper presented at: MIT Sloan Analytics Conference, Boston MA. 2016.

15. Dower C, Rafehi A, Weber J, Mohamad R. An enhanced metric of injury risk utilizing artificial intelligence. Paper presented at: MIT Sloan Analytics Conference, Boston, MA. 2018. https://www.alerteds.com/wp-content/uploads/2016/02/MITSSAC2018-An-enhanced-metric-of-injury-risk-utilizing-Artificial-Intelligence.pdf

16. Lysholm J, Wiklander J. Injuries in runners. *Am J Sports Med*. 1987;15(2):168–171. PubMed ID: 3578639 doi:10.1177/036354658701500213

17. Van Der Does HTD, Brink MS, Otter RTA, Visscher C, Lemmink KAPM. Injury risk is increased by changes in perceived recovery of team sport players. *Clin J Sport Med*. 2017;27(1):46–51. PubMed ID: 26945309 doi:10.1097/JSM.0000000000000306

18. Rogalski B, Dawson B, Heasman J, Gabbett TJ. Training and game loads and injury risk in elite Australian footballers. *J Sci Med Sport*. 2013;16(6):499–503. PubMed ID: 23333045 doi:10.1016/j.jsams.2012.12.004

19. Nielsen RØ, Parner ET, Nohr EA, Sørensen H, Lind M, Rasmussen S. Excessive progression in weekly running distance and risk of running-related injuries: an association which varies according to type of injury. *J Orthop Sports Phys Ther*. 2014;44(10):739–747. PubMed ID: 25155475 doi:10.2519/jospt.2014.5164

20. Chen T, Guestrin C. Xgboost: A scalable tree boosting system. Paper presented at: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining; 2016:785–794. doi:10.1145/2939672.2939785.

21. Nielsen D. *Tree Boosting With XGBoost-Why Does XGBoost Win" Every" Machine Learning Competition?* [Master's thesis]. Trondheim, Norway: Norwegian University of Science and Technology; 2016.

22. Alsahaf A, Azzopardi G, Ducro B, Hanenberg E, Veerkamp R, Petkov N. Estimation of muscle scores of live pigs using a kinect camera. *IEEE Access*. 2019;7:52238–52245. doi:10.1109/ACCESS.2019.2910986.

23. Farrugia S, Ellul J, Azzopardi G. Detection of illicit accounts over the Ethereum blockchain. *Expert Syst Appl*. 2020;150:113318. doi:10.1016/j.eswa.2020.113318

24. Developers XGBoost. XGBoost v0.9 Documentation. 2016. https://xgboost.readthedocs.io/en/release_0.90/. Accessed July 26, 2019.

25. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. *Eur Heart J*. 2014;35(29):1925–1931. PubMed ID: 24898551 doi:10.1093/eurheartj/ehu207

26. Menaspa P, Sassi A, Impellizzeri FM. Aerobic fitness variables do not predict the professional career of young cyclists. *Med Sci Sports Exerc*. 2010;42(4):805–812. PubMed ID: 19952851 doi:10.1249/mss.0b013e3181ba99bc

27. Hulin BT, Gabbett TJ, Lawson DW, et al. The acute:chronic workload ratio predicts injury: high chronic workload may decrease injury risk in elite rugby league players. *Br J Sports Med*. 2016;50(4):231–236. PubMed ID: 26511006 doi:10.1136/bjsports-2015-094817

28. Heidari J, Beckmann J, Bertollo M, et al. Multidimensional monitoring of recovery status and implications for performance. *Int J Sports Physiol Perform*. 2019;14(1):2–8. doi:10.1123/ijspp.2017-0669

29. Op De Beéck T, Jaspers A, Brink MS, et al. Predicting future perceived wellness in professional soccer: the role of preceding load and wellness. *Int J Sports Physiol Perform*. 2019;14(8):1074–1080. doi:10.1123/ijspp.2017-0864

30. Lieli RP, Hsu YC. Using the area under an estimated ROC curve to test the adequacy of binary predictors. *J Nonparametr Stat*. 2019;31(1):100–130. doi:10.1080/10485252.2018.1537440

31. Memmert D, Lemmink KAPM, Sampaio J. Current approaches to tactical performance analyses in soccer using position data. *Sports Med*. 2017;47(1):1–10. PubMed ID: 27251334 doi:10.1007/s40279-016-0562-5