Personalized machine learning approach to injury monitoring in elite volleyball players

Leeuw, A.W. de; Zwaard, S. van der; Baar, R. van; Knobbe, A.J.

European Journal of Sport Science

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tejs20

Personalized machine learning approach to injury monitoring in elite volleyball players

Arie-Willem de Leeuw, Stephan van der Zwaard, Rick van Baar & Arno Knobbe

To cite this article: Arie-Willem de Leeuw, Stephan van der Zwaard, Rick van Baar & Arno Knobbe (2022) Personalized machine learning approach to injury monitoring in elite volleyball players, European Journal of Sport Science, 22:4, 511-520, DOI: 10.1080/17461391.2021.1887369

To link to this article: https://doi.org/10.1080/17461391.2021.1887369

Published online: 25 Feb 2021.

EUROPEAN JOURNAL OF SPORT SCIENCE
2022, VOL. 22, NO. 4, 511–520
https://doi.org/10.1080/17461391.2021.1887369

Routledge
Taylor & Francis Group

OPEN ACCESS Check for updates

Personalized machine learning approach to injury monitoring in elite volleyball players

Arie-Willem de Leeuw[a], Stephan van der Zwaard[a,b], Rick van Baar[c] and Arno Knobbe[a]

[a]Leiden Institute of Advanced Computer Science (LIACS), Leiden University, Leiden, The Netherlands; [b]Department of Human Movement Sciences, Faculty of Behavioral and Movement Sciences, Amsterdam Movement Sciences, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands; [c]The Dutch Volleyball Federation (Nevobo), Utrecht, the Netherlands

ABSTRACT
We implemented a machine learning approach to investigate individual indicators of training load and wellness that may predict the emergence or development of overuse injuries in professional volleyball. In this retrospective study, we collected data of 14 elite volleyball players (mean ± SD age: 27 ± 3 years, weight: 90.5 ± 6.3 kg, height: 1.97 ± 0.07 m) during 24 weeks of the 2018 international season. Physical load was tracked by manually logging the performed physical activities and by capturing the jump load using wearable devices. On a daily basis, the athletes answered questions about their wellness, and overuse complaints were monitored via the Oslo Sports Trauma Research Center (OSTRC) questionnaire. Based on training load and wellness indicators, we identified subgroups of days with increased injury risk for each volleyball player using the machine learning technique Subgroup Discovery. For most players and facets of overuse injuries (such as *reduced sports participation*), we have identified personalized training load and wellness variables that are significantly related to overuse issues. We demonstrate that the emergence and development of overuse injuries can be better understood using daily monitoring, taking into account interactions between training load and wellness indicators, and by applying a personalized approach.

Highlights
- With detailed, athlete-specific monitoring of overuse complaints and training load, practical insights in the development of overuse injuries can be obtained in a player-specific fashion contributing to injury prevention in sports.
- A multi-dimensional and personalized approach that includes interactions between training load variables significantly increases the understanding of overuse issues on a personal basis.
- Jump load is an important predictor for overuse injuries in volleyball.

KEYWORDS
Injury; training load; volleyball; machine learning; personalization

Introduction

In many sports, elite athletes need to maximize training volumes to improve their physical capacities. However, if they do not include sufficient recovery, there is an increased risk of overtraining and overuse injuries (Bacon & Mauger, 2017; Martínez-Silván, Díaz-Ocejo, & Murray, 2017; Wilson et al., 2010). Most studies in this area focus on the incidence of or time-loss due to injuries (Anderson et al., 2003; Brink et al., 2010; Hulin et al., 2014). However, important additional information on overuse injuries is provided by the severity of experienced complaints, which can be assessed by the Oslo Sports Trauma Research Center (OSTRC) overuse injury questionnaire (Clarsen, Myklebust, & Bahr, 2013). Moreover, it is important to monitor injury incidences and severity of complaints frequently, preferably once a day, to prevent that short-lasting complaints or injuries will be missed (Clarsen et al., 2013) or that the onset of an overuse injury is not detected in time (Hespanhol Junior et al., 2015).

Besides frequent data collection and a comprehensive analysis of the severity of complaints, detailed monitoring of the athlete's training load (Eckard et al., 2018; Jones, Griffiths, & Mellalieu, 2017), responses to this training load and other lifestyle factors (Buchheit et al., 2013; Hooper & Mackinnon, 1995) is also essential for a better understanding of the many factors involved in the

CONTACT Arie-Willem de Leeuw a.de.leeuw@liacs.leidenuniv.nl Leiden Institute of Advanced Computer Science (LIACS), Leiden University, Niels Bohrweg 1, 2333 CA, Leiden, The Netherlands

development of injuries, and individual differences therein. Analysing all these dimensions involved in injury development allows investigation of the most important predictors, as well as the interactions between predictors. In sports, the benefits of analysing multiple variables or predictors and their (potentially non-linear) interactions has been shown for the physical characteristics of athletes (López-Valenciano et al., 2018; Rommers et al., 2020; Van der Zwaard et al., 2019) or characteristics of training load (Jaspers et al., 2018; Knobbe et al., 2017). Additionally, the training load is affected by individual characteristics (Jones et al., 2017), and therefore we need a new and personalized approach to identify athlete-specific indicators for injuries (Bartlett et al., 2017).

Part of our approach is the use of a machine learning technique known as Subgroup Discovery (Klösgen & Zytkow, 2002; Novak, Lavrac, & Webb, 2009). While the collected data allows the application of other techniques, we will argue that Subgroup Discovery has a number of attractive properties that justify its employment. First, we aim to discover interpretable and actionable dependencies in the data, such that we require a method that selects the most important predictors and describes succinctly how these predictors influence the onset of overuse injuries. Second, we expect overuse injuries to occur in non-linear fashion, especially where certain load indicators surpass the (individual) thresholds of the athlete in question. As such, common linear modelling methods may not be the best choice. Subgroup Discovery is a good fit because it meets the two requirements mentioned, while not being as uninterpretable and demanding as other machine learning methods that have gained popularity of late (e.g. neural networks). Finally, with Subgroup Discovery, a model can be learned for each individual athlete and predict the development of overuse injuries for each individual athlete (given that sufficient data is available) and thus provide personalized advice. Personalization is especially of interest in the context of team sports, where the coaches and staff need to accommodate for the various positions or roles in the team as well as differences in the physiological strengths and weaknesses of every athlete.

The aim of this study is to elucidate player-specific relationships between training load, wellness and overuse complaints by applying the machine learning technique Subgroup Discovery, in the domain of elite volleyball. As in volleyball most overuse issues are related to the knee (Kilic et al., 2017), we hypothesize that jump load is an important predictor for (symptoms of) overuse injuries. Moreover, we demonstrate that our findings reveal personalized predictors of complaints that can straightforwardly be interpreted by the coach, and incorporated in the training schedule of each individual player, in order to reduce their personal risk of overuse injuries.

Methods

Subjects

Fourteen elite male volleyball players volunteered to participate in this study (mean ± SD age: 27 ± 3 years, weight: 90.5 ± 6.3 kg, height: 1.97 ± 0.07 m). All subjects competed on the international level, represented the same country and provided written informed consent. One player was excluded due to insufficient data entry. Three players were excluded from the machine learning analyses, because the absence of any complaints related to overuse issues hindered the identification of player-specific relationships between training load, wellness and overuse complaints. Note that even though data of only 10 players was involved, our analysis is player-specific, and each player provided a considerable data set of sufficient statistical power.

In our machine learning analyses, the sample size, i.e. the number of days on which a player is monitored, was 94 ± 18 data points per athlete (mean ± SD). In hypothesis testing, a power analysis is used to justify the sample size. Here, the statistical validation is not related to the total number of participants, but to the number of data points per volleyball player. For a data collection of a given player, we determine the probability that a finding truly exists, or is a spurious finding caused by insufficient data. The details of this procedure are discussed later on in this section.

Experimental design

Training sessions and matches were monitored during 24 weeks of the 2018 international volleyball season for national teams. The season started with a 5-week preparation phase, followed by 3 weeks of competition. Hereafter, the players had holidays for 3 weeks. In this period, the players had no training activities and therefore no data was collected during their vacations. Subsequently, there were 10 weeks of training, followed by another 3 weeks of competition. Excluding the three weeks without training activities, we find that on a weekly basis, the players participate in 6.1 ± 2.4 training sessions and train for 13.8 ± 5.0 h.

Data collection

Injuries, illness and severity of complaints

Players completed the OSTRC overuse injury questionnaire (Clarsen et al., 2014). Hereto, subjects answered questions regarding their overuse problems in the following categories:

Q1: **Difficulty participating** in normal training and competition.

Q2: **Reduced training volume**.

Q3: **Affected performance**.

Q4: **Experienced symptoms/complaints**.

Since we have elite athletes, we opted for daily monitoring with slightly modified questions to only reflect on complaints in the previous 24 h (Clarsen et al., 2020). Participants were instructed to answer the OSTRC questionnaire every morning before breakfast, except during holidays. To facilitate comparison, the four scores were normalized to four severity scores that range from 0 (completely healthy) to 100 (maximal disturbance due to complaints).

Although the discretization of the OSTRC variables is a limitation, the OSTRC questionnaire is considered a valid method for investigating overuse injuries in sports (Gallagher et al., 2017). The method has been tested and validated in high-level athletes from different countries and sports (Charlton et al., 2017 Clarsen et al., 2014; Ekman et al., 2015; Hirschmüller et al., 2017; Jorgensen et al., 2016; Nagano et al., 2019), and captures the full development of overuse injuries. There is a high internal consistency (average Cronbach's $\alpha > 0.9$), good test-retest reliability (intraclass correlation coefficient 0.62–0.91) and high construct validity (Ekman et al., 2015; Hirschmüller et al., 2017; Jorgensen et al., 2016).

Perceived wellness

Data on subjective ratings of wellness (e.g. sleep, fatigue) was obtained individually and included questions about fatigue, sleep quality, number of hours slept, general muscle soreness and mood. Questions were answered each morning before breakfast (except for holidays) and were rated on a 10-point Likert scale ranging from 1 (very bad) to 10 (excellent). Volleyball players were familiarized with the questionnaire before the onset of the study.

Training load

For the strength training sessions, we obtained the recorded number of repetitions, the number of sets, the applied weight and finished exercises. The weight applied in the exercises was reported both in absolute kilograms and relative to the individual's one-repetition maximum (1-RM).

For the volleyball-specific training sessions and matches, we obtained the number of jumps and jump heights of each jump using the G-VERT (Mayfonk Inc., Fort Lauderdale, FL, USA), which was firmly secured to the trunk, near the center of mass, using an elastic band (Charlton et al., 2017; MacDonald et al., 2017). In previous studies (Charlton et al., 2017; Skazalski et al., 2018), it has been shown that the G-VERT reports jumps above 15 cm and detects these jumps with 99% accuracy. Moreover, the sensor has a high inter-device reliability and showed excellent agreement with other reference systems (Charlton et al., 2017; Skazalski et al., 2018).

The internal training load was captured for strength training and volleyball-specific sessions by the rating of perceived exertion (RPE) using the CR10-scale (Borg, Hassmen, & Lagerstrom, 1987). RPE scores were multiplied by session duration to get the session loads, which were used to calculate training loads (summation of session loads), monotony (day-to-day variation in training load) and strain (overall stress) (Haddad et al., 2017).

Data analysis

In our analysis, the outcome variables were the four answers to the OSTRC questions, which we analyse one at a time. Predictor variables were those on internal and external training load and perceived wellness. Our personalized approach, which assumes different players have different "weak spots", calls for a larger collection of predictors than is usually considered.

A player's physical well-being on a specific day will be influenced by the training activities over the preceding days. Therefore, we construct predictors that are aggregate functions of training load and wellness in the preceding days. We consider three different time windows (i.e. the preceding 7, 14 and 28 days), to distinguish short, mid and long-term effects and as aggregate functions, we consider the mean, standard deviation, first quartile, third quartile, and sum. In principle, we could also have used other distribution-related measures such as the median or median absolute deviation. However, having too many predictors, especially when highly collinear, can drastically increase the number of hypotheses tested, and may thus negatively affect the statistical significance of the findings. Therefore, we focus on a moderate set of functions that are interpretable and we believe capture the most relevant information, avoiding excessive numbers of predictors.

For the data on jump performance specifically, 72 potential predictors were constructed. We consider the number of jumps, jump heights and categorized the number of jumps with *low* (<50 cm), *average* (between 50 and 65 cm) and *high* (higher than 65 cm) jump height. For strength training, we divided the exercises into *full body*, *lower body* and *upper body exercises*. The 81 constructed predictors focus on the weight of the corresponding exercises. Moreover, 48 predictors were constructed from the perceived wellness of the players. As muscle soreness can be directly related to overuse issues, this facet of the perceived wellness is not included as predictor. Additionally, we included 27 predictors that concern the training load, monotony and strain (Haddad et al., 2017), for all training sessions together as well as for strength training or volleyball-specific training sessions separately. Finally, 9 predictors captured the frequency of training sessions (three session types (*all*, *strength*, *ball*) in the three time windows).

Machine learning: Subgroup Discovery

We cannot afford to build a multi-variate regression model involving all predictors, due to the risk of overfitting. Instead, we employ a machine learning method that considers the influence of one, or only a few predictors on the dependent variable of choice, and reports only the most important predictors. Our method of choice is the supervised technique of *Subgroup Discovery*. A beneficial side-effect of this method is that the results are easy to interpret and therefore can be put into practice straightforwardly (de Leeuw, Meerhoff, & Knobbe, 2018; Knobbe et al., 2017).

In Subgroup Discovery, we start by choosing a specific outcome variable whose distribution we want to understand, i.e. the *target variable*. The goal is then to detect subgroups for which the distribution of the target variable is different from that of the entire data. The obtained subgroups are characterized by one or multiple conditions for the predictors (i.e. predictor values are above or below specific thresholds) and are therefore capable of capturing non-linear effects.

Differences in the distribution of the target variable between the subgroup and the individual athlete's entire data set are quantified by a *quality measure*, which captures the magnitude of the dependency between the involved predictors and the target variable. In this study, we use the *Explained Variance* (EV) (Knobbe et al., 2017), a quality measure inspired by the R^2 employed in linear regression. The EV is a numerical value in the range from 0 to 1, where a larger value indicates a stronger dependency.

Subgroup Discovery is designed for the analysis of data with many (potentially correlated) variables. When considering large numbers of variables, and combinations thereof, we face the multiple comparison problem (Hochberg & Tamhane, 1987), i.e. we risk finding spurious results only as a consequence of testing many hypotheses. A common way to circumvent this risk in Subgroup Discovery is through the so-called *Distribution of False Discoveries* (Duivesteijn & Knobbe, 2011) (DFD). This method repeats the Subgroup Discovery process on (say) a thousand randomized versions of the data, where the relation between predictors and targets is deliberately broken (through swap-randomization). Such "false" runs should not produce any significant results, so the spurious scores that do turn up on randomized data are indicative of the false discoveries produced by multiple hypothesis testing. A 5% significance threshold on the EV is now produced from the DFD for each volleyball player and OSTRC question combination separately, which can then be used on the real data to determine what level of dependence constitutes a significant finding.

Subgroup Discovery implementation

We use the OSTRC questions as outcome variables. Therefore, we have four different ordinary targets, i.e. discretized indicators that characterize symptoms of overuse. We will apply Subgroup Discovery for each target and each player. We start by finding subgroups characterized by a single predictor, after which we look for more complex effects, that is, subgroups that are described by two or more predictors. We only consider subgroups encompassing between 5% and 95% of the data, to prevent obtaining too specific results.

We primarily consider the subgroups with the largest EV. We will use these results for three analyses. First, we compare findings for each of the four OSTRC questions. Second, we compare the results for subgroups involving either one or multiple predictor variables. Third, we investigate to what extent results differ per player.

Statistical analysis

Subgroups identified by Subgroup Discovery were considered to be significant if their explained variance exceeded the 5% significance threshold produced from the DFD. To compare the explained variance between the found subgroups that are characterized by single or multiple predictors, we use a paired sample *t*-test. We consider the difference between both to be significant if $p < 0.05$. Moreover, we use Cohen's *d* to determine effect sizes, including 95% confidence intervals (95% CI).

Results

Severity of complaints

Professional volleyball players completed a total of 1112 questionnaires, and in 313 cases (28.1%), the players reported overuse complaints, i.e. severity score > 0 for

Table 1. The average severity score for each OSTRC question, per player. Note that some players do not report any reduced training volume or affected performance. Also, players differ in their susceptibility to injury.

Player	Difficulty participating	Reduced training volume	Affected performance	Experienced symptoms/ complaints
P01	9.3%	3.7%	8.0%	9.3%
P02	6.0%	4.3%	6.3%	5.7%
P03	8.3%	8.7%	9.3%	8.7%
P04	12.7%	12.7%	14.0%	14.0%
P05	19.0%	18.3%	19.7%	20.3%
P06	8.7%	1.3%	6.3%	6.3%
P07	12.3%	0.0%	0.0%	12.7%
P08	2.0%	0.3%	3.7%	2.0%
P09	6.3%	0.0%	1.3%	15.0%
P10	7.0%	6.0%	7.0%	9.0%

at least one of the four questions. In 25 entries (2.2%), athletes reported substantial complaints (severity score ≥ 50) for *affected performance* (Q3) or *reduced training volume* (Q2) (Clarsen et al., 2014). Average severity scores for each player and OSTRC question are shown in Table 1.

Development of overuse complaints

Figure 1 illustrates the development of an overuse complaint in player P01 over the course of 9 days. Note how the scores for each question change from day to day, which is an advantage of using a daily instead of weekly monitoring frequency of overuse issues. Next, Figure 1 reveals that the development is expressed differently by the four questions in terms of timing and severity. In this example, the order of importance is: *experienced symptoms/complaints* (Q4), *affected performance* (Q3), *difficulty participating* (Q1), *reduced training volume* (Q2). Our results show that the course of overuse complaints can be accurately described with daily monitoring of the OSTRC questionnaire, capturing the four dimensions with a full range and resolution of the scales.

Machine learning with single predictors

The results of the Subgroup Discovery analysis with single predictors are displayed in Table 2 and Figure 2. In the table, we show the most important predictors (i.e. conditions described by a training load indicator, time window and threshold) of the subgroups for all participants, for question Q1 (for the sake of brevity). The subgroups and their severity scores were presented for each of the OSTRC questions in Figure 2. Note that no significant subgroups could be detected in eight cases (see Figure 2B–D), three times because the volleyball player entered no complaints for the respective OSTRC question in all of his entries. Our findings show that subgroups with high severity scores could be detected based on single predictors for each of the OSTRC questions and for most of the volleyball players.

As a demonstration of our results, consider the answers of player P02 to question Q1 (*difficulty participating*). From Figure 2(A), we find that the subgroup of

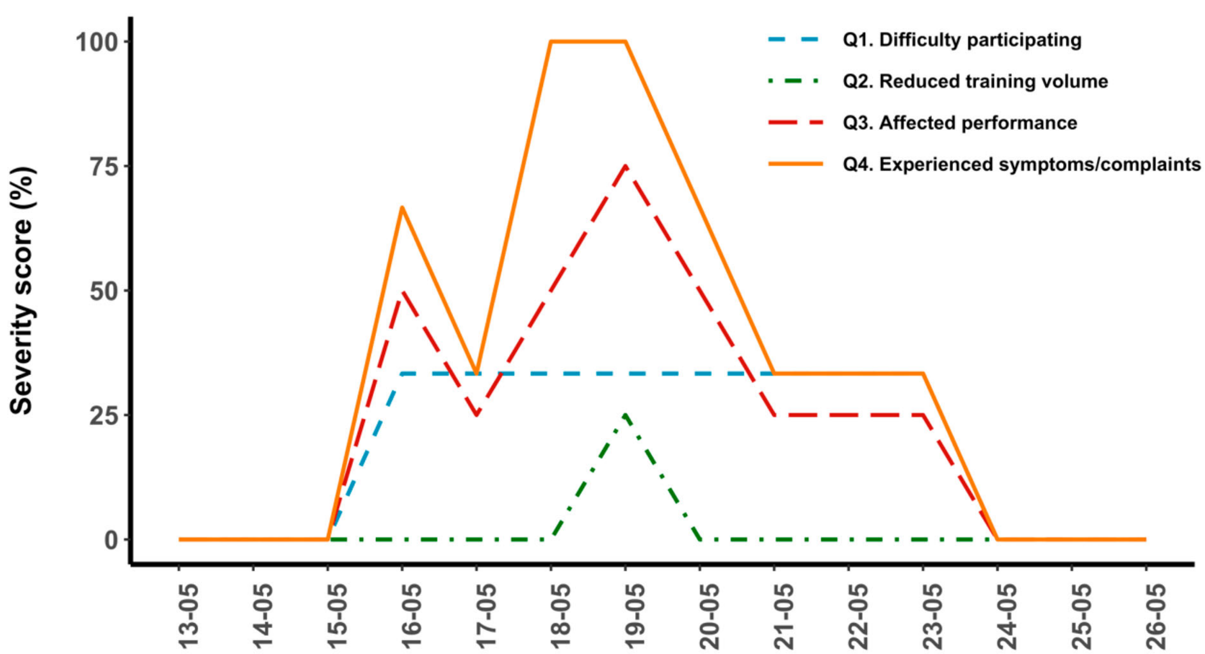

Figure 1. The course of the severity scores for the four facets of the OSTRC questionnaire for player P01, for a selected range of dates. For each component, the severity scores 0 and 100 indicate no complaints or maximal hindrance, respectively.

Table 2. Overview of the most important subgroups that are characterized by a single condition for the OSTRC question that indicates whether player had *difficulty participating* (Q1). For all players, we show the relevant time window and the condition on the predictor variable of the subgroup with the largest explained variance. The result for player P02 has been described as an example in the Results section.

Player	Window	Condition	Subgroup size
P01	14 days	Stand. dev. number of jumps ≥ 86.8[a]	28%
P02	14 days	Total number of jumps ≥ 196[a]	6%
P03	28 days	Stand. dev. daily number of jumps ≤ 22.2[a]	28%
P04	14 days	Average jump height ≥ 54.7 cm	24%
P05	28 days	Third quartile number of jumps ≥ 65.75	42%
P06	28 days	First quartile of daily mood scores ≥ 8	40%
P07	28 days	Stand. dev. daily number of high jumps ≥ 1.70	38%
P08	14 days	Average daily sleep duration ≤ 7.11 h	16%
P09	14 days	Average jump height ≤ 48.3 cm	12%
P10	14 days	Stand. dev. weight percentage upper body exercises ≥ 0.08	10%

[a]The best subgroup is also described by other jump-specific characteristics.

P02 has an average severity score of 44.3%, which is significantly higher than his total average severity score of 6% ($p \leq 0.05$).[1] The size of the subgroup (% of the data points), as well as the predictor that explains most of the variance in the severity scores, can be found in Table 2. For P02, this suggests that high severity scores on the OSTRC question for *difficulty participating* could be avoided by performing not more than 196 jumps over a span of 14 days.

Machine learning with multiple predictors

In addition to single predictors, we also performed Subgroup Discovery to detect subgroups using multiple predictors. For example, consider player P08 and the responses to the *difficulty participating* question. For single predictors, the largest value of the explained variance, i.e. 0.344, was obtained if the average hours slept in the previous 14 days is less than 7.11. If, in addition to this condition, the standard deviation of his mood scores in the preceding 14 days is less than 1.03, the explained

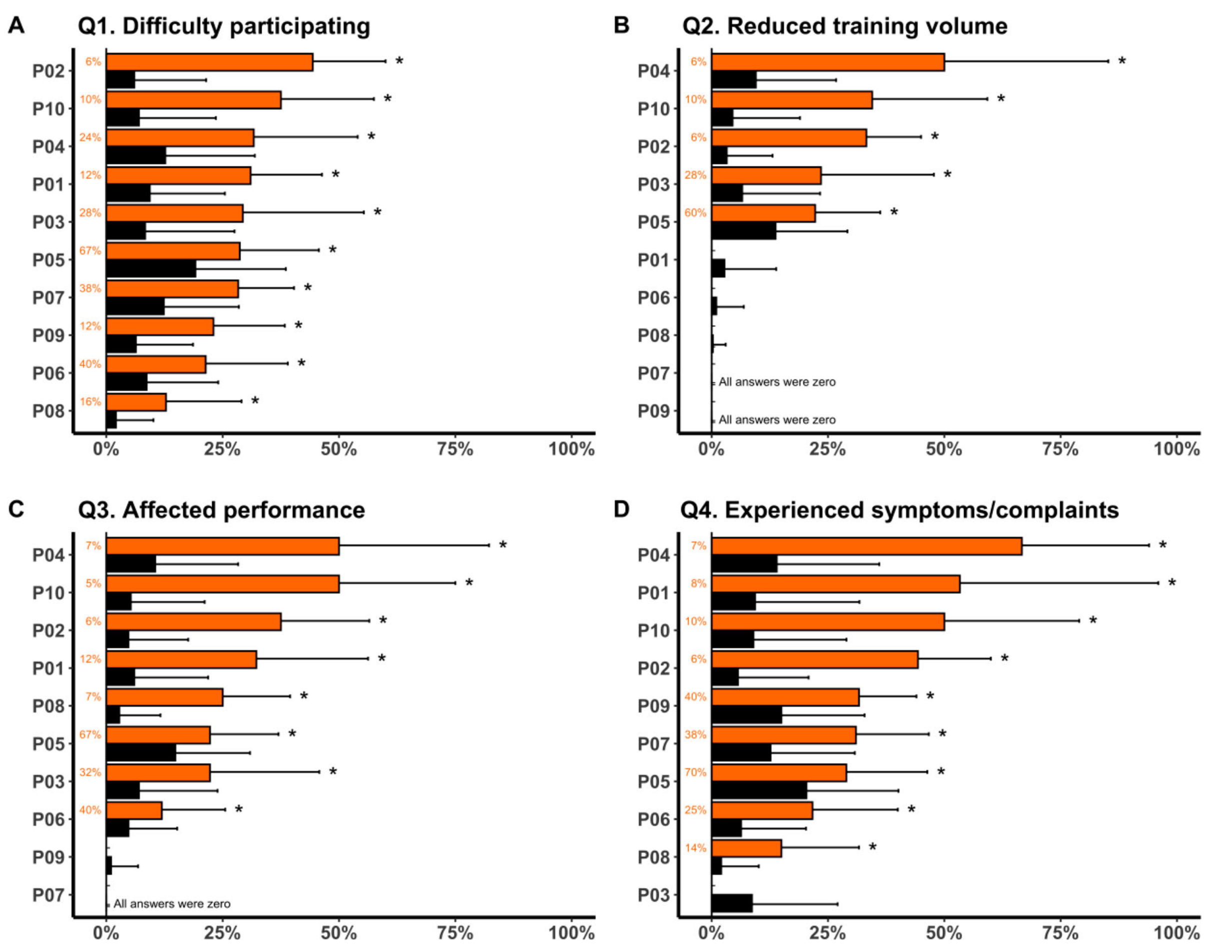

Figure 2. Overall (black) and subgroup (orange) mean severity scores are shown for each individual athlete. Severity scores are displayed for each OSTRC question and for the strongest subgroup using a single predictor. The bar reflects the mean severity score, the line shows the standard deviation and the number at the bottom of the bar equals the relative size of the subgroup.

variance increases to 0.695. Thus, by monitoring a combination of the hours slept and mood, almost twice as much variance can be explained.

We have found that more of the variance in the severity scores can be explained with two predictors instead of one for the OSTRC questions *experienced symptoms/complaints* ($p = 0.004$; $d = 1.52$, 95% CI: [0.73,3.13]), *difficulty participating* ($p < 0.001$; $d = 1.74$, 95% CI: [1.13,3.26]) and *reduced training volume* ($p = 0.03$; $d = 0.28$, 95% CI: [0.08, 0.63]). For *affected performance*, the difference was not significant ($p = 0.30$; $d = -0.29$, 95% CI: [−0.96, 0.21]). If we consider three instead of two predictors, we can only explain more variance for *difficulty participating* ($p = 0.01$; $d = 0.64$, 95% CI: [0.26, 1.36]). For more than three predictors, there is no significant evidence that more of the variance in the responses to the OSTRC questions can be obtained.

Jump load

In our machine learning analyses, we observed that training load indicators that described jump load were important predictors of subgroups with high overuse injury severity scores. For example, in Table 1, we demonstrate for single predictors and *difficulty participating*, that for 70% of the professional volleyball players, the predictors are related to the jump load. Therefore, our findings confirmed the importance of jump load in overuse injuries in volleyball.

Discussion

In this section, we discuss the results of our approach and also elaborate on the practical implications of this study.

Overuse issues

We have collected daily data on the development and emergence of overuse issues via the OSTRC questionnaire. Based on daily data, in 28% of the questionnaires, the players reported complaints and the percentage of substantial complaints in our study is only 2.2%. However, considering weekly scores (maximum scores during the week), players report a complaint in 49% of the weeks and the percentage of substantial complaints is 10%.

Compared to previous studies, we have found that our percentage of weekly complaints is similar to the high prevalence of overuse complaints in swimming (Nagano et al., 2019) and is higher than the 22% that is observed on average for injuries in handball, orienteering, volleyball and tennis (Ekman et al., 2015) or the 31% that is reported in team sports (Clarsen et al., 2014). Therefore, with daily monitoring, more small complaints are reported. The percentage of substantial complaints in our study is comparable to other studies based on weekly monitoring (Clarsen et al., 2014; Ekman et al., 2015; Nagano et al., 2019).

Predictors

We have investigated the dependence between training load and overuse complaints by simultaneously considering multiple predictors related to training load and perceived wellness. Following the suggestions of previous studies (Coyne et al., 2018; Rabello L et al., 2019), we have applied a personalized approach.

We have found individual differences in the most relevant descriptor of the training load and the time window that is considered. This suggests that every athlete has his own predictor for monitoring overuse issues. For example, player P01, had *difficulty participating* in training when the standard deviation in the number of jumps in the previous 14 days was larger than 86.8. On the other hand, players P04 and P08 had most difficulties participating if the average jump height in the previous 14 days was larger than 54.7 cm or the average number of slept hours in the past 14 days was less than 7.11, respectively.

Previous studies focused on the effect of single predictors and have found increased risks on overuse complaints for higher ratios between acute and chronic loads (Bowen et al., 2017; Hulin et al., 2014), rapid changes in weekly training load (Gabbett, 2016) and a high jump count (Visnes & Bahr, 2013). Our analyses confirm the role of jump load, but identifies more detailed and athlete-specific aspects of this load. For example, for *difficulty participating*, the total number of jumps is the most important predictor for player P02. On the other hand, for player P03 and P07, the variation in the number of (high) jumps is most important and finally the jump height is most relevant for P04 and P09.

Advantages of Subgroup Discovery

Our machine learning approach of Subgroup Discovery has several advantages. First, we have the possibility of finding player-specific relationships that we would have missed, had we grouped all players in one analysis. A prime example is our libero, who had *difficulty participating* in training when the number of jumps in the previous 14 days was larger than 196. Although 196 jumps doesn't seem excessive, note that a libero mainly stands and attacks the ball from a reaching position (no jumping). Therefore, this number of jumps is high for this type of player, which illustrates the strength of our player-specific analysis. Moreover, by applying Subgroup Discovery for each player separately, we have found conditions on predictors where each player had different severity scores than usual. Therefore, the individual bias, i.e. the possible differences of the average responses for different players, does not affect our analysis.

The second advantage of our approach is that the results are interpretable and can therefore straightforwardly be put into practice. Consider player P08, who had *difficulty participating* if the average hours slept in the previous 14 days is less than 7.11. Therefore, it is very important that the staff monitors his sleep, for example while travelling across different time zones.

Supporting the high practicality of our results, the staff already fine-tuned their training regime based on our findings. For example, we found that one of the players experienced symptoms or complaints if he performed upper body strength training in the past week with on average more than 63 kilograms. When discussing this with the technical staff, they were already aware that this player had complaints after completing strength training sessions with heavy weights. However, the staff only had a rough estimate what weight was too much for the upper body strength exercises, and our finding proved to be a valuable addition by specifying this threshold.

The third advantage of using Subgroup Discovery is the straightforward possibility to also consider non-linear interactions between predictors. We have demonstrated that for 3 of the 4 facets of injury monitoring, significantly more variance in the answers can be explained if the interactions are included in the analyses. Moreover, we have found that with interactions between four or more predictors, there is no significant evidence that more of the variance can be explained.

Limitations of our approach

We acknowledge that there are certain limitations to our approach. First, overuse issues occurring on a certain day might affect subsequent training activities. However, in this study only in 1.3% of the cases a player had to reduce his training activities to at least a moderate extent. Moreover, our predictors are aggregates over 7, 14 or 28 days and the most important predictors concern windows of 2 or 4 weeks, as for example can be seen in Table 2. Thus, although these time-dependent confounding issues are small, it would be worthwhile to investigate these effects by for example using Markov chains.

Second, we made no explicit distinction between illness and injuries per body part. We distinguished between general symptoms/complaints and issues that affect training participation or performance, but it would be interesting to consider for example knee or shoulder issues separately (Clarsen et al., 2015).

Practical implications

Although our player-specific approach hinders a generalization of concrete results to a completely different population, we have obtained relevant results for practitioners in volleyball and other sports.

First, we have demonstrated that a frequent, daily monitoring of overuse complaints decreases the prevalence of substantial issues. Second, we have found that jump load predictors are important for overuse issues in volleyball. Depending on individual characteristics and position in the team, different aspects, such as jump count and height, are most relevant. Third, we have shown that with detailed, athlete-specific monitoring of overuse complaints and training load, practical insights in the development of overuse injuries can be obtained in a player-specific fashion. This identifies the strengths and weaknesses of each athlete and enables direct application of the found results, which contributes to injury prevention in sports using a personalized advice.

Conclusion

We applied a machine learning approach to obtain player-specific relationships between overuse complaints and training load and wellness indicators. We demonstrated that the emergence and development of overuse issues can be better understood if the monitoring occurs on a daily basis, the interactions between multiple training load variables are taken into account and a personalized approach is used. The results indicate that tracking the jump load is an important step towards the prevention of overuse issues in volleyball. As our personalized findings are easy to interpret, these can be used to minimize injury risks when designing training schemes for each individual player.

Acknowledgements

We are grateful to the athletes and coaching staff of the volleyball team for their participation and collaboration.

Disclosure statement

No potential conflict of interest was reported by the author(s).

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.

Contributors

All authors contributed to the development and design of the study. RvB was responsible for the data collection, SvdZ and AWdL analysed the data. All authors interpreted the results, provided revisions and contributed to the final manuscript.

Ethics approval

The study was approved by the ethics committee of the Vrije Universiteit Amsterdam (VCWE-2019-118) and was conducted in agreement with the Helsinki Declaration.

Note

1. Note that the significance of the result implies that the sample size (94) for this specific player is sufficient.

References

Anderson, L., Triplett-McBride, T., Foster, C., et al. (2003). Impact of training patterns on incidence of illness and injury during a women's collegiate basketball season. *Journal of Strength and Conditioning Research*, *17*(4), 734–738.

Bacon, C. S., & Mauger, A. R. (2017). Prediction of overuse injuries in professional u18-u21 footballers using metrics of training distance and intensity. *Journal of Strength and Conditioning Research*, *31*(11), 3067–3076.

Bartlett, J. D., O'Connor, F., Pitchford, N., et al. (2017). Relationships between internal and external training load in team sport athletes: Evidence for an individualised approach. *International Journal of Sports Physiology and Performance*, *12*(2), 230–234.

Borg, G., Hassmen, P., & Lagerstrom, M. (1987). Perceived exertion related to heart rate and blood lactate during arm and leg exercise. *European Journal of Applied Physiology and Occupational Physiology*, *56*(6), 679–685.

Bowen, L., Gross, A. S., Gimpel, M., et al. (2017). Accumulated workloads and the acute: Chronic workload ratio relate to injury risk in elite youth football players. *British Journal of Sports Medicine*, *51*(5), 452–459. doi:10.1136/bjsports-2015-095820

Brink, M. S., Visscher, C., Arends, S., et al. (2010). Monitoring stress and recovery: New insights for the prevention of injuries and illnesses in elite youth soccer players. *British Journal of Sports Medicine*, *44*(11), 809–815.

Buchheit, M., Racinais, S., Bilsborough, J. C., et al. (2013). Monitoring fitness, fatigue and running performance during a pre-season training camp in elite football players. *Journal of Science and Medicine in Sport*, *16*(6), 550–555. doi:10.1016/j.jsams.2012.12.003. [published Online First: 2013/01/22].

Charlton, P. C., Kenneally-Dabrowski, C., Sheppard, J., et al. (2017). A simple method for quantifying jump loads in volleyball athletes. *Journal of Science and Medicine in Sport*, *20*(3), 241–245.

Clarsen, B., Bahr, R., Heymans, M. W., et al. (2015). The prevalence and impact of overuse injuries in five Norwegian sports: Application of a new surveillance method. *Scandinavian Journal of Medicine & Science in Sports*, *25*(3), 323–330. doi:10.1111/sms.12223. [published Online First: 2014/04/02].

Clarsen, B., Bahr, R., Myklebust, G., et al. (2020). Improved reporting of overuse injuries and health problems in sport: An update of the Oslo Sport Trauma Research Center questionnaires. *British Journal of Sports Medicine*, *54*(7), 390–396. doi:10.1136/bjsports-2019-101337

Clarsen, B., Myklebust, G., & Bahr, R. (2013). Development and validation of a new method for the registration of overuse injuries in sports injury epidemiology. *British Journal of Sports Medicine*, *47*(8), 495–502.

Clarsen, B., Rønsen, O., Myklebust, G., et al. (2014). The Oslo Sports Trauma Research Center questionnaire on health problems: A new approach to prospective monitoring of illness and injury in elite athletes. *British Journal of Sports Medicine*, *48*(9), 754–760.

Coyne, J. O. C., Haff, G. G., Coutss, A. J., et al. (2018). The current state of subjective training load monitoring – a practical perspective and call to action. *Sports Medicine Open*, *4*(1), 58–67.

de Leeuw, A.-W., Meerhoff, L. A., & Knobbe, A. J. (2018). Effects of pacing properties on performance in long-distance running. *Big Data*, *6*(4), 248–261.

Duivesteijn, W., & Knobbe, A. J. (2011). Exploiting false discoveries - statistical validation of patterns and quality measures in subgroup discovery. *ICDM, International Conference on Data Mining*, 151–160.

Eckard, T. G., Padua, D. A., Hearn, D. W., et al. (2018). The relationship between training load and injury in athletes: A systematic review. *Sports Medicine (Auckland, NZ)*, *48*, 1929–1961.

Ekman, E., Frohm, A., Ek, P., et al. (2015). Swedish translation and validation of a web-based questionnaire for registration of overuse problems. *Scandinavian Journal of Medicine & Science in Sports*, *25*(1), 104–109.

Gabbett, T. J. (2016). The training-injury prevention paradox: Should athletes be training smarter and harder? *British Journal of Sports Medicine*, *50*(5), 273–280.

Gallagher, J., Needleman, I., Ashley, P., et al. (2017). Self-reported outcome measures of the impact of injury and illness on athlete performance: A systematic review. *Sports Medicine (Auckland, NZ)*, *47*(7), 1335–1348. doi:10.1007/s40279-016-0651-5

Haddad M, Stylianides G, Djaoui L, et al. Session-RPE method for training load monitoring: Validity, ecological usefulness, and influencing factors. *Frontiers in Neuroscience* 2017;11(612)

Hespanhol Junior, L. C., Barboza, S. D., van Mechelen, W., et al. (2015). Measuring sports injuries on the pitch: A guide to use in practice. *Brazilian Journal of Physical Therapy*, *19*(5), 369–380. doi:10.1590/bjpt-rbf.2014.0110. [published Online First: 2015/11/06].

Hirschmüller, A., Steffen, K., Fassbender, K., et al. (2017). German translation and content validation of the OSTRC questionnaire on overuse injuries and health problems. *British Journal of Sports Medicine*, *51*(4), 260–263.

Hochberg, Y., & Tamhane, A. (1987). *Multiple comparison procedures*. New York: Wiley.

Hooper, S. L., & Mackinnon, L. T. (1995). Monitoring overtraining in athletes. Recommendations. Sports Medicine (Auckland. *NZ*, *20*(5), 321–327. doi:10.2165/00007256-199520050-00003

Hulin, B. T., Gabbett, T. J., Blanch, P., et al. (2014). Spikes in acute workload are associated with increased injury risk in elite cricket fast bowlers. *British Journal of Sports Medicine*, *48* (8), 708–712.

Jaspers, A., de Béeck T, O., Brink, M. S., et al. (2018). Relationships between the external and internal training load in professional soccer: What can we learn from machine learning? *International Journal of Sports Physiology and Performance*, *13*(5), 625–630.

Jones, C. M., Griffiths, P. C., & Mellalieu, S. D. (2017). Training load and fatigue marker associations with injury and illness: A systematic review of longitudinal studies. *Sports Medicine (Auckland, NZ)*, *47*, 943–974.

Jorgensen, J. E., Rathleff, C. R., Rathleff, M. S., et al. (2016). Danish translation and validation of the Oslo Sports Trauma Research Centre questionnaires on overuse injuries and health problems. *Scandinavian Journal of Medicine & Science in Sports*, *26*(12), 1391–1397.

Kilic, O., Maas, M., Verhagen, E., et al. (2017). Incidence, aetiology and prevention of musculoskeletal injuries in volleyball: A systematic review of the literature. *European Journal of Sport Science*, *17*(6), 765–793.

Klösgen, W., & Zytkow, J. M. (2002). *Handbook of data mining and knowledge discovery*. New York, NY: Oxford University Press, Inc.

Knobbe, A. J., Orie, J., Hofman, N., et al. (2017). Sport analytics for professional speed skating. *Data Mining and Knowledge Discovery*, *31*(6), 1872–1902.

López-Valenciano, A., Ayala, F., Puerta, J. M., et al. (2018). A Preventive model for muscle injuries: A Novel approach based on learning algorithms. *Medicine & Science in Sports & Exercise*, *50*(5), 915–927. doi:10.1249/mss.0000000000001535. published Online First: 2017/12/29].

MacDonald, K., Bahr, R., Baltich, J., et al. (2017). Validation of an inertial measurement unit for the measurement of jump count and height. *Physical Therapy in Sport*, *25*, 15–19.

Martínez-Silván, D., Díaz-Ocejo, J., & Murray, A. (2017). Predictive indicators of overuse injuries in adolescent endurance athletes. *International Journal of Sports Physiology and Performance*, *12*(Suppl 2), S2153–S2S56.

Nagano Y, Kobayashi-Yamakawa K, Higashihara A, et al. Japanese translation and modification of the Oslo Sports Trauma Research Centre overuse injury questionnaire to evaluate overuse injuries in female college swimmers. *PLos One*, 2019;*14*(4), e0215352.

Novak, P., Lavrac, N., & Webb, G. I. (2009). Supervised descriptive rule discovery: A unifying survey of contrast set, emerging pattern and subgroup mining. *Journal of Machine Learning Research*, *10*, 377–403.

Rabello L, M., Zwerver, J., Stewart, R. E., et al. (2019). Patellar tendon structure responds to load over a 7-week preseason in elite male volleyball players. *Scandinavian Journal of Medicine & Science in Sports*, *29*(7), 992–999.

Rommers, N., Rossler, R., Verhagen, E., et al. (2020). A machine learning approach to assess injury risk in elite youth football players. *Medicine & Science in Sports & Exercise*, doi:10.1249/mss.0000000000002305. [published Online First: 2020/02/23].

Skazalski, C., Whiteley, R., Hansen, C., et al. (2018). A valid and reliable method to measure jump-specific training and competition load in elite volleyball players. *Scandinavian Journal of Medicine & Science in Sports*, *28*(5), 1578–1585.

Van der Zwaard, S., de Ruiter, C. J., Jaspers, R. T., et al. (2019). Anthropometric clusters of competitive cyclists and their sprint and endurance performance. *Frontiers in Physiology*, *10*, 1276.

Visnes, H., & Bahr, R. (2013). Training volume and body composition as risk factors for developing jumper's knee among young elite volleyball players. *Scandinavian Journal of Medicine & Science in Sports*, *23*(5), 607–613.

Wilson, F., Gissane, C., Gormley, J., et al. (2010). A 12-month prospective cohort study of injury in international rowers. *British Journal of Sports Medicine*, *44*(3), 207–214.