SYSTEMATIC REVIEW   **Open Access**

Current Approaches to the Use of Artificial Intelligence for Injury Risk Assessment and Performance Prediction in Team Sports: a Systematic Review

João Gustavo Claudino[1,2*], Daniel de Oliveira Capanema[3], Thiago Vieira de Souza[2], Julio Cerca Serrão[1], Adriano C. Machado Pereira[4] and George P. Nassis[5,6]

Abstract

Background: The application of artificial intelligence (AI) opens an interesting perspective for predicting injury risk and performance in team sports. A better understanding of the techniques of AI employed and of the sports that are using AI is clearly warranted. The purpose of this study is to identify which AI approaches have been applied to investigate sport performance and injury risk and to find out which AI techniques each sport has been using.

Methods: Systematic searches through the PubMed, Scopus, and Web of Science online databases were conducted for articles reporting AI techniques or methods applied to team sports athletes.

Results: Fifty-eight studies were included in the review with 11 AI techniques or methods being applied in 12 team sports. Pooled sample consisted of 6456 participants (97% male, 25 ± 8 years old; 3% female, 21 ± 10 years old) with 76% of them being professional athletes. The AI techniques or methods most frequently used were artificial neural networks, decision tree classifier, support vector machine, and Markov process with good performance metrics for all of them. Soccer, basketball, handball, and volleyball were the team sports with more applications of AI.

Conclusions: The results of this review suggest a prevalent application of AI methods in team sports based on the number of published studies. The current state of development in the area proposes a promising future with regard to AI use in team sports. Further evaluation research based on prospective methods is warranted to establish the predictive performance of specific AI techniques and methods.

Keywords: Artificial neural networks, Machine learning, Technology, SportsTech, Innovation, Analytics, Soccer, Basketball, Handball, Volleyball

Key Points

- AI techniques or methods currently having been employed for sporting performance prediction are artificial neural network, decision tree classifier, Markov process, and support vector machine in sports such as basketball, soccer, and volleyball.

- For injury risk assessment, the artificial neural network, decision tree classifier, and support vector machine have been used in soccer, basketball, American football, Australian football, and handball.

- Application of AI methods in team sports has the potential to grow further given continued development of the field and implementation of evaluation research in sports practice to establish the predictive performance of each specific technique/method.

* Correspondence: claudinojgo@usp.br
[1]University of São Paulo, School of Physical Education and Sport - Laboratory of Biomechanics, Av. Prof. Mello de Morais, 65 – Cidade Universitária, São Paulo, São Paulo 05508-030, Brazil
[2]Research and Development Department, LOAD CONTROL, Contagem, Minas Gerais, Brazil
Full list of author information is available at the end of the article

Background

Artificial intelligence (AI) techniques and methods have attracted considerable attention in the information industry and in society as a whole, due to the large amount of data and the imminent need to transform this data into useful knowledge and practical solutions [1–3]. However, the effective use of data in some areas is still under development, as is the case in sports. As in most other areas of society, increasing volume of data has been gathered in all kinds of sports, and automated data analysis became an important and fast developing field. Careful analyses of these large data sets can enhance our knowledge in sport sciences while at the same time assist in the decision-making of the practitioners who work on the optimization of training and competition strategies [4, 5].

Data science has emerged as a strategical area to exploit knowledge in sports science aiming to fill some gaps left by traditional statistical methods. As a hybrid knowledge area, data science is more than the combination of statistics and computer science as it requires training in how to weave statistical and computational techniques into a larger framework, problem by problem, and to address discipline-specific questions [6]. A holistic view of data science requires an understanding of the context of data, appreciating the responsibilities involved in using private and public data, and clearly communicating what a dataset can and cannot tell us about the real world [6], in our case, in the sports world. Based on learning models, the algorithms can be tuned and be optimized in order to produce better results for supporting decisions and provide applied knowledge to athletes and sport professionals. These algorithms are applied as supervised learning (e.g., classification and regression) and unsupervised learning (e.g., clustering). The supervised learning requires input and output data to develop a predictive model whereas the unsupervised learning is based on input data only [7]. Data science is an emerging area in both industry and academic ways, leading to more evidence-based decision-making across many walks of life, including social networking services, streaming services, health care, manufacturing, education, financial modeling, policing, and marketing [8–11]. Also, in our area, linking science and technology to increase sports efficiency has been touted as a path with a promising future. For this to happen, the work directed toward the innovation, introduction, and improvement of processes performed by research and development (R&D) departments in the world's largest technology companies has been also suggested in the sport field. Because of its fast-moving environment, sport professionals combine data (e.g., physical, technical and tactical) with their expert opinion to inform decisions on the players [12].

However, success in team sport, from the sport science and medicine staff perspective, is to use evidence-based knowledge in an effective manner to develop the decision-making process for injury risk reduction and athletes' performance optimization [13]. How can the adoption of the research findings and innovations be improved hence improving the compliance of players and coaches/managers with injury prevention and performance enhancement programs? To answer this question, we must accept that players' needs come first and that players and coaches are the main actors. Therefore, we need to add value to them, offering solutions that really affect their daily lives [13].

One key issue in the sports industry is the possibility to predict injury risk and performance. Historically, the ability of the coaching staff to prescribe training to achieve optimal athletic performance with low risk of injury can be attributed to many years of personal experience. However, modern approaches aiming in adopting scientific methods for the effective development of optimal training programs are warranted [14]. The application of contemporary statistical approaches from AI open an interesting perspective for dealing with injury prevention and for improving the performance models [15, 16].

Therefore, the understanding of the state-of-the-art of AI techniques or methods applied to team sport warrants investigation. With this in mind, the purpose of this study is to give an overview of the current state of the application of AI in team sports. In particular, our review is aiming in answering the following questions: (1) Which AI approaches have been applied to studies investigating injury risk and sport performance in team sports? (2) What are the team sports that have been using AI techniques for predicting injury risk and athletes' performance?

Methods

Procedures

The review methodology adopted the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [17]. The selection process and data extraction methods were completed by JGC, DOC, and TVS. The quality appraisal was completed by the same authors.

Search Strategy

Three electronic databases (PubMed, Web of Science, and Scopus) were systematically searched up to May 2018. The command line ("machine learning" OR "predictive modeling" OR "injury prediction" OR "learning algorithms" OR "data mining" OR "naïve bayes" OR "logistic regression" OR "random forest" OR "support vector machine" OR "neural network" OR "deep learning" OR "artificial intelligence" OR "extreme learning machines" OR "data

science" OR "knowledge discovery" OR "injury forecasting" OR "injury detection" OR "decision trees" OR "business intelligence") AND ("team sport" OR "team sports" OR "sport" OR "sports" OR "individual sports" OR "individual sport" OR "athlete" OR "athletes") AND ("monitoring" OR "load" OR "training load" OR "controlling" OR "control" OR "load control" OR "regulating" OR "regulation" OR "managing" OR "management" OR "improvement" OR "improve" OR "optimizing" OR "optimize" OR "enhance" OR "enhancement" OR "performance" OR "reduce" OR "reducing" OR "decrease" OR "decreasing" OR "injury risk" OR "injury prevention") was used during the electronic search and applied for the last 5 years to find out the current state of the application of AI in team sports.

Eligibility Criteria and Selection Process

The first author (JGC) reviewed and identified the titles and abstracts based on the following inclusion criteria with two other authors (DOC and TVS) double checked. If some doubt arose, the fourth author (ACMP) was involved for the final decision. The inclusion criteria were:

1. The study was written in English
2. The study was published as a full-text, original research paper in a peer-reviewed journal
3. Data was reported just for team sport or individual sport athletes and split into distinct groups
4. The participants were competitive athletes (defined as olympic, international, professional, semi-professional, national, youth elite or division I collegiate)
5. The AI techniques or algorithms should be described and tested

Quality Assessment

The quality of all studies was evaluated using evaluation criteria (Table 1) described by Saw et al. [18]. Scores were allocated based on how well each criterion was met, assuming a maximum possible score of 7 (low risk of bias). Studies with a risk of bias score of 4 or less were considered poor and were excluded. Once the studies to be included were selected, we performed a

review on checking reference lists [19] to identify additional peer-reviewed studies.

Classifying the Main Research AI Technique or Method

Initially, we considered only the main learning AI technique or method [1–3] for this systematic review. Then, we divided our methodology into two sections, the main method, which corresponds to the main AI technique or method used, and a complete method, which describes all the techniques used in the adopted model by a research paper. This organization aims to facilitate the classification of each paper and compare it with others. Furthermore, we registered the model evaluation metrics used to assess the performance of the technique/method employed. When more than one AI technique or method was evaluated in the same study, we decided to highlight the best model in this paper, that is, the model with the best performance.
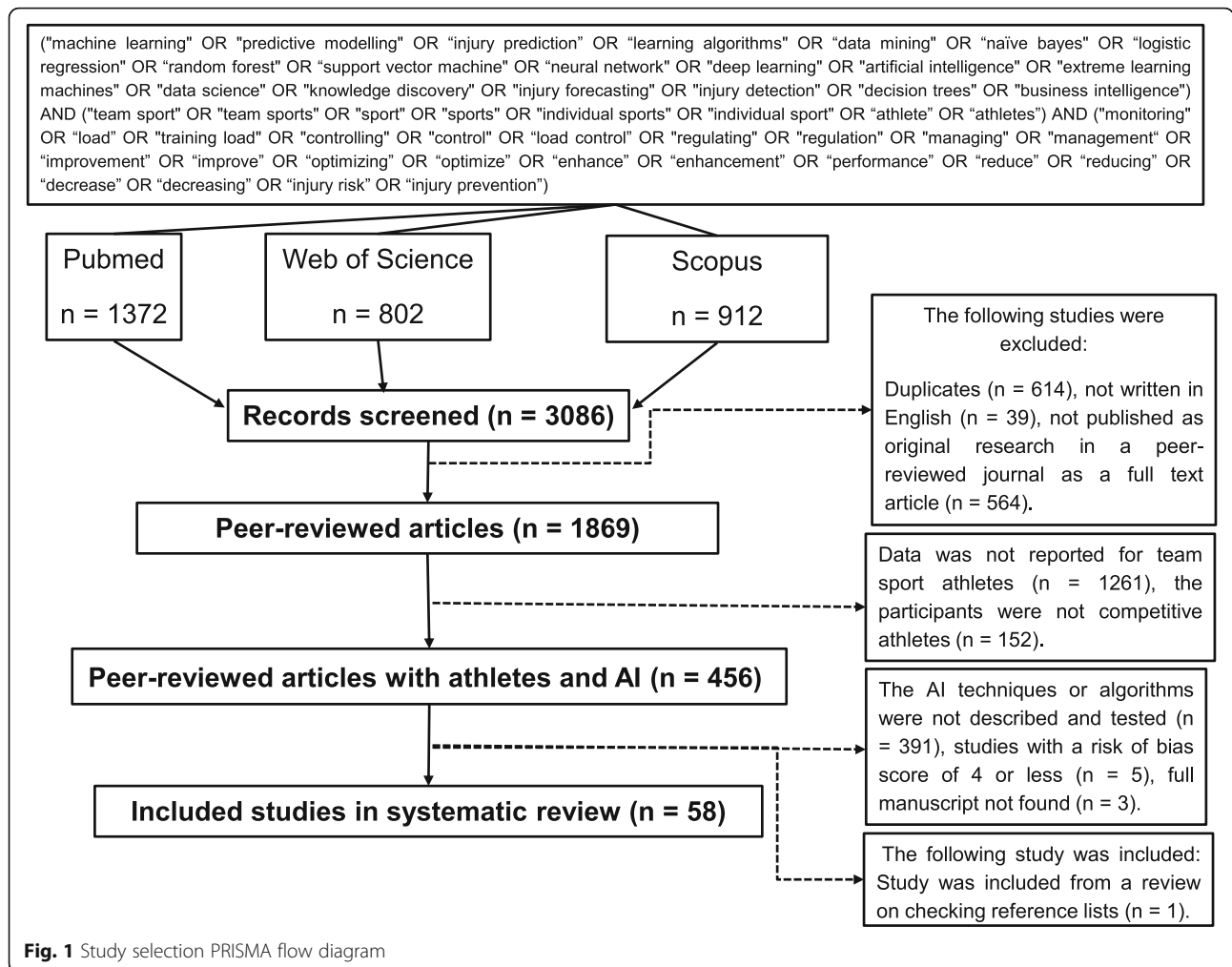
Results

The initial search returned 3086 articles (for details see Fig. 1). After the removal of duplicated articles ($n = 614$), a total of 2472 studies were retained for full-text screening. Following eligibility assessment, 5 studies with a risk of bias score of 4 or less were considered poor and were excluded (for details see Additional file 1: Table S1). The remaining studies were evaluated between 6 and 7 points in terms of quality. We were unable to find the full manuscript for 3 research studies which were excluded from further analysis. During the revision of the reference lists, 1 study found meeting all the criteria and was included in the analysis. Thus, 58 studies in total were included in this systematic review [15, 20–76].

Characteristics of the Studies

The pooled sample was 6456 participants (i.e., the sum of the subjects of all studies that reported the sample size, i.e., 48% of the studies) with the vast majority being male (97% male, 25 ± 8 years old; 3% female, 21 ± 10 years old). Seventy-six percent of them were professional, 10% youth academy, 9% collegiate, 5% semi-professional and national team U18, and 2% retired professional players. Twenty-six percent of the sample were

Table 1 Risk of bias assessment criteria

Criteria		Definition	Scoring		
			0	1	2
A	Peer-reviewed	Study published in peer-reviewed journal	No	Yes	–
B	Real-world approach	The approach was performed with real results/data of the athletes	No	Yes	–
C	Population defined	Age, gender, sport, and level was described	No	Partly	Yes
D	Experimental design	Experimental design of the study period was described and replicable	No	Partly	Yes
E	Artificial intelligence	The artificial intelligence approaches/techniques were described	No	Yes	–

("machine learning" OR "predictive modelling" OR "injury prediction" OR "learning algorithms" OR "data mining" OR "naïve bayes" OR "logistic regression" OR "random forest" OR "support vector machine" OR "neural network" OR "deep learning" OR "artificial intelligence" OR "extreme learning machines" OR "data science" OR "knowledge discovery" OR "injury forecasting" OR "injury detection" OR "decision trees" OR "business intelligence") AND ("team sport" OR "team sports" OR "sport" OR "sports" OR "individual sports" OR "individual sport" OR "athlete" OR "athletes") AND ("monitoring" OR "load" OR "training load" OR "controlling" OR "control" OR "load control" OR "regulating" OR "regulation" OR "managing" OR "management" OR "improvement" OR "improve" OR "optimizing" OR "optimize" OR "enhance" OR "enhancement" OR "performance" OR "reduce" OR "reducing" OR "decrease" OR "decreasing" OR "injury risk" OR "injury prevention")

Pubmed	Web of Science	Scopus
n = 1372	n = 802	n = 912

Records screened (n = 3086)

The following studies were excluded:

Duplicates (n = 614), not written in English (n = 39), not published as original research in a peer-reviewed journal as a full text article (n = 564).

Peer-reviewed articles (n = 1869)

Data was not reported for team sport athletes (n = 1261), the participants were not competitive athletes (n = 152).

Peer-reviewed articles with athletes and AI (n = 456)

The AI techniques or algorithms were not described and tested (n = 391), studies with a risk of bias score of 4 or less (n = 5), full manuscript not found (n = 3).

Included studies in systematic review (n = 58)

The following study was included: Study was included from a review on checking reference lists (n = 1).

Fig. 1 Study selection PRISMA flow diagram

soccer players, 22% basketball, 10% handball and volleyball, 9% Australian football, 9% baseball, 7% American football, 5% ice hockey, 3% rugby, and 2% each were beach volleyball, cricket, and field hockey players.

Main Artificial Intelligence (AI) Technique or Method in Team Sports

Eleven AI techniques or methods (Table 2) [1–3] were found in 12 team sports. Approximately two thirds (n = 43) of the AI studies were related to sporting performance (74%) whereas 15 studies were related to injury risk (26%). With regard to the injury risk assessment studies, 27% of them were related to training load, 13% to concussion, screening, and, training process/knee injury causes, and 7% each were on ground reaction force pattern, heart defect detection, monitoring based on wearable sensors, psychosocial stress factors, and ulnar collateral ligament reconstruction approaches. With regard to the performance studies, 88% of them were related to technical and tactical analysis, 5% were on physical, technical, and tactical analysis, and 2% each were on match attendance,

psychological dynamics of cooperative teamwork, and prediction based on heart rate measures.

In the last 5 years, the main AI technique or method used for injury risk assessment and sporting performance prediction was artificial neural network (10% of the injury risk and 26% of the sports performance prediction studies, reporting its use). For injury risk assessment studies, the decision tree classifier and support vector machine (5%) were the next mostly used techniques and methods. In the performance prediction domain, the decision tree classifier (17%), Markov process, and support vector machine (9%) were the most frequently AI techniques/methods used. AI for injury risk assessment was applied to soccer (12% of the studies), basketball, American football, Australian football, and handball (3%) whereas basketball (19%), soccer (14%), and volleyball (9%) were the sports which mostly used performance prediction algorithms. The AI techniques or methods with the best model evaluation metrics were indicated to be applied (for details see Additional file 2: Table S2). Furthermore, 11 studies [22, 25, 26, 29, 37, 48, 49, 52,

Table 2 AI techniques or methods descriptions

AI	Description
Absolute shrinkage and selection operator	In statistics and machine learning, least absolute shrinkage and selection operator (LASSO) is a regression analysis method that performs both variable selection and regularization in order to enhance the prediction accuracy and interpretability of the statistical model it produces. It was originally introduced in the geophysics literature in 1986 but was independently rediscovered and popularized by Robert Tibshirani in 1996; he coined the term and provided further insights into the observed performance. LASSO was originally formulated for least squares models. This simple case provides a vast scenario regarding the behavior of the estimator, including its relationship to ridge regression, best subset selection, besides the connections between LASSO coefficient estimates and the so-called soft thresholding. It also reveals that (as standard linear regression) the coefficient estimates need not be unique if covariates are collinear. Although originally defined for least squares, LASSO regularization is easily extended to a wide variety of statistical models including generalized lin ear models, generalized estimating equations, proportional hazards models, and M-estimators, in a straightfor ward fashion.
Artificial neural network	An artificial neural network (ANN) is a computational model based on the structure and functions of biological neural networks. Information that flows through the network affects the structure of the ANN because a neural network changes—or learns, in a sense—based on inputs and outputs. ANNs are considered nonlinear statistical data modeling tools whereby the complex relationships between inputs and outputs are modeled or patterns are found.
Bayesian logistic	In statistics, the logistic model (or logit model) is a widely used statistical model that, in its basic form, uses a logistic function to model a binary dependent variable; many more complex extensions exist. In regression analysis, logistic regression (or logit regression) means estimating the parameters of a logistic model; it is a form of binomial regression. Logistic regression is a type of regression analysis used for predicting the outcome of a categorical (a variable that can take on a limited number of categories) dependent variable based on one or more predictor variables. The probabilities describing the possible outcome of a single trial are modeled, as a function of explanatory variables, using a logistic function. Logistic regression measures the relationship between a categorical dependent variable and usually a continuous independent variable (or several), by converting the dependent variable to probability scores. Logistic regression can be binomial or multinomial. In the multinomial type, there exist different models to perform the regression, such as the Bayesian method.
Bayesian networks	Bayesian networks are a type of probabilistic graphical model that uses Bayesian inference for probability computations. Bayesian networks aim to model conditional dependence and, therefore, causation, by representing conditional dependence by edges in a directed graph. Through these relationships, one can efficiently conduct the inference of the random variables in the graph by using factors.
Decision tree classifier	A decision tree classifier (DTC) is a decision support tool that uses a tree-like graph or model of decisions and their possible consequences, including chance event outcomes, resource costs, and utility. It is one way to dis play an algorithm that only contains conditional control statements. Decision trees are commonly used in oper ations research, specifically in decision analysis, to help identify a strategy most likely to reach a goal, but are also a popular tool in machine learning. Decision tree learning uses a decision tree (as a predictive model) to go from observations about an item (represented in the branches) to conclusions about the item target value (represented in the leaves). It is one of the predictive modeling approaches used in statistics, data mining, and machine learning. Tree models whereby the target variable can take a discrete set of values are called classifica tion trees; in these tree structures, leaves represent class labels and branches represent conjunctions of features that lead to those class labels.
Fuzzy clustering	Fuzzy clustering is an alternative method to conventional or hard clustering algorithms, which partitions data containing similar subjects. Fuzzy clustering contrasts with hard clustering by its nonlinear nature and discipline of flexibility in grouping massive data. It provides more accurate and close-to-nature solutions for partitions and herein implies more possibilities of solutions for decision-making. In the specific matter of computation, fuzzy clustering has its roots in fuzzy logic and indicates the likelihood or degrees of one data point belonging to more than one group.
K-means clustering	K-means clustering is a type of unsupervised learning, used when there is unlabeled data (i.e., data without defined categories or groups). The goal of this algorithm is to find groups in the data, with the number of groups represented by the variable K. The algorithm works iteratively to assign each data point to one of the K groups based on the features provided. Data points are clustered based on feature similarity. The results of the K-means clustering algorithm are the centroids of the K clusters, which can be used to label new data, and labels for the training data (each data point is assigned to a single cluster). Rather than defining groups before looking at the data, clustering allows finding and analyzing the groups that have formed organically. Each centroid of a cluster is a collection of feature values which define the resulting groups. Examining the centroid feature weights can be used to qualitatively interpret what kind of group each cluster represents.
K-nearest neighbor	K-nearest neighbor (k-NN) is a type of instance-based learning, or lazy learning, in which the function is only approximated locally and all computation is deferred until classification. The k-NN algorithm is among the sim plest of all machine learning algorithms. Both for classification and regression, a useful technique can be used to assign weight to the contributions of the neighbors, so that the nearer neighbors contribute more to the average than the more distant ones. For example, a common weighting scheme consists in giving each neigh bor a weight of $1/d$, where d is the distance to the neighbor. The neighbors are taken from a set of objects for which the class (for k-NN classification) or the object property value (for k-NN regression) is known. This can

Table 2 AI techniques or methods descriptions *(Continued)*

AI	Description
	be thought of as the training set for the algorithm, although no explicit training step is required. A peculiarity of the *k*-NN algorithm is that it is sensitive to the local structure of the data. The algorithm is not to be confused with *k*-means, another popular machine learning technique.
Markov process	A Markov process is a random process indexed by time, and with the property that the future is independent of the past. The present. Markov processes, named for Andrei Markov, are among the most important of all random processes. In a sense, they are the stochastic analogs of differential equations and recurrence relations, which are certainly among the most important deterministic processes.
Support vector machine	Support vector machine (SVM) is a discriminative classifier formally defined by a separating hyperplane. In other words, given the labeled training data (supervised learning), the algorithm outputs an optimal hyperplane which categorizes new examples. In a two-dimensional space, this hyperplane is a line dividing a plane into two parts in which each class lies on either side. The learning of the hyperplane in linear SVM is done by transforming the problem using some linear algebra. This is where the kernel plays its role.
Support vector machine + decision tree classifier	The idea of combining SVM and DTC is to provide a hybrid approach, which attempts to embed SVM within a decision tree algorithm as a decision tree pre-pruning method and resulting into a more accurate and efficient hybrid classifier.

53, 55, 69] did not report the evaluation metrics specific for the model. However, the authors of the latter studies recommended the application of AI techniques or methods tested in each manuscript. Only one study did not recommend the use of the tested AI [70] (for details see Additional file 2: Table S2; Figs. 2 and 3). AI techniques or methods had better performance metrics than traditional statistical methods for predicting injury risk and athletes' performance in 8 studies [31, 35, 38, 54, 57, 67, 70, 71] that made this comparison.

Discussion

The main purpose of this study is to provide an overview of the current state of the application of AI in assessing the injury risk and predicting performance in team sports athletes. The AI techniques or methods with the greatest potential were artificial neural networks, decision tree classifier, Markov process, and support vector machine.

Lapham and Bartlett [77]'s study was one of the first studies using AI in the analysis of sports performance. In 1995, they demonstrated that the increasing role of computers in the decision-making process, through the use of AI techniques, would be a potentially rewarding future direction for the discipline and already pointed out the artificial neural networks successes in experiments, along with their potential benefits. Another study [78] published over 20 years ago reported the use of AI, with decision tree classifier and Bayesian classification, in the diagnosing of sport injuries. Naive Bayesian classifier with the fuzzy discretization of numerical attributes was superior to other methods and was estimated as the most appropriate for practical use [78]. However, this method is not so popular nowadays, whereas the decision tree classifier has been mostly used so far.

McCall et al. [12] recommended that professional soccer teams included R&D into their daily activities to reduce injuries and to optimize performance. Most of the participants in the studies included in our paper were soccer players (26%), and this could be due to the fact that most of the studies were in soccer. In this particular sport, AI was applied to "training load" [71, 73], "training process/knee injury causes" [63], "heart defect detection" [64], "ground reaction force pattern" [65], "psychosocial stress factors" [69], and "screening" [15]. In summary, the AI techniques or methods applied to predict the injury risk ($n = 7$ studies) in the sample are composed of collegiate (14%) and professional (86%) soccer players and were distributed as following: 43% were on artificial neural network [63–65], 29% on decision tree classifier [15, 71], and 14% on each of Bayesian logistic regression [69] and least absolute shrinkage and selection operator [73] (for details see Additional file 2: Table S2). Regarding the soccer performance domain, AI was applied to "technical and tactical analysis" [30, 36, 44, 45, 48, 50], "match attendance" [31], as well as on "psychological dynamics of cooperative teamwork" [35]. The AI techniques or methods used to predict sporting performance ($n = 8$ studies) in youth academy (13%), semi-professional (13%), and professional (75%) soccer players were as follows: 25% each of artificial neural network [30, 31], Bayesian network [35, 36] and decision tree classifier [44, 45], and 13% each of fuzzy clustering [48] and *K*-means clustering [50]. A great challenge for the future is the application of AI models to integrate all related variables for injury risk assessment and performance prediction.

The second sport in AI application was basketball. Most of the studies on performance were in the "technical and tactical analysis" area [20, 22–25, 41–43, 55, 62], whereas one study was on the "physical, technical, and tactical analysis" [21]. Their focus justifies the recognized importance of physical, technical, and tactical

Fig. 2 AI for predicting injury risk in various sports

analysis for basketball success [79, 80]. On a separate note, one study described a system for the automatic detection and tracking of the ball trajectory during a free throw based on algorithms [81]. According to the authors of that paper, coaches using this system will be able to monitor the trajectory of the ball and the parameters of the free throw, a fact that will help them correct the technique and hence improve the athletic performance of basketball players [81]. Our analysis showed that the AI techniques or methods used to predict sporting performance (*n* = 11 studies) were with youth academy (9%), semi-professional (9%), and professional (82%) basketball players. The methods reported were as follows:

on each artificial neural network (55%) [20–25], on decision tree classifier (27%) [41–43], and 9% each of Markov process [55] and support vector machine + decision tree classifier [62]. For injury risk assessment, the studies were based on "training process/knee injury causes" [66] and "heart defect detection" [64]. The prediction of knee injury risk is essential due to the high costs of medical treatment [82] and to the high rates of medically disqualifying injuries among young athletes [83]. Heart defect detection for basketball players is also highly desired because sudden cardiac death risk among males was estimated at more than 10 times higher than that in the overall athletic population of the

Fig. 3 AI for predicting sporting performance in various sports

National Collegiate Athletic Association (NCAA) Division 1 (i.e., 1 in 5,200 vs. 1 in 53,703 athletes per year) [84, 85]. For injury risk prediction, the most frequently used AI technique or method (*n* = 2 studies) applied in collegiate and professional basketball players was artificial neural network [64, 66].

AI application into handball and volleyball were well reported. In terms of injury risk assessment, ground reaction force patterns is a standard method of investigation in sports medicine and biomechanics [65] that is also related to knee injuries in multidirectional sports, such as handball and volleyball [86, 87]. These injuries

may occur during activities, such as jump landing, cutting, and pivoting [86, 87]. Furthermore, injury prediction with AI technique or method based on "screening" [15] was successfully applied to handball players. This approach may assist in identifying the injury risk with higher probability compared to the use of isolated screening tests [88, 89]. The AI techniques or methods (*n* = 2 studies) used to predict injury risk in professional handball (100%) and volleyball (50%) players were artificial neural network [65] and decision tree classifier [15]. For handball performance, all the studies were on the "technical and tactical analysis" area with youth academy

based studies using artificial neural network ($n = 4$ studies) [26–29]. For volleyball performance, most of the studies were on the "technical and tactical analysis" area too [33, 52, 53, 61] and only one, with professional players, was on "physical, technical, and tactical analysis" [51]. AI techniques or methods mostly used for performance prediction ($n = 5$ studies) were the Markov process (40%) [52, 53], artificial neural network (20%) [33], *K*-nearest neighbor (20%) [51], and support vector machine (20%) [61]. The areas of physical, technical, and tactical analysis are of high importance to athletic performance in both sports [90, 91], and coaches/players may now be able to use the tools of AI for better decision-making.

Training load management is a concern in Australian football [92–94]. Along this line, AI methods and techniques have been used to predict the risk of injury through the "training load" [67] and "screening" [70] in that sport. Interestingly, there is only one study where the authors did not recommend the supervised learning techniques based on "screening" [70]. In that study, the use of eccentric hamstring strength, age, and previous hamstring strain injury [70] was not able to predict the risk of injury. In this latter study, the Bayesian network was applied in professional Australian football players with the area under the curve (AUC) of 54% being classified as a poor performance metric [70]. On the other hand, when the artificial neural network based on "training load" was used with a sample of professional athletes of the same sport, it was recommended by the authors [67]. The prediction of performance in Australian football was based on heart rate [38] and "technical and tactical analysis" [39, 40] data with the use of decision tree classifier ($n = 3$ studies in total). The authors concluded that AI could add benefits and, hence, was recommended in Australian football [38–40].

Injuries cost over 1 billion dollars to Major League Baseball teams, and thus, preventing them is of high priority in that sport too [95]. There is a predominance of upper extremity injuries in baseball players, the surgical procedure most closely associated with baseball being the reconstruction of the ulnar collateral ligament of the elbow [96]. So far, AI has been used to predict the injury risk after ulnar collateral ligament reconstruction in professional baseball players [76]. With regard to performance improvement in baseball [95], we only found AI approaches in the "technical and tactical analysis" area [32, 37, 57, 58]. The AI techniques or methods used to predict sporting performance ($n = 4$ studies) in professional baseball players were support vector machine [57, 58], artificial neural network [32], and Bayesian network [37].

Data show that American football players who sustained concussions may experience significant salary reductions and perform worse after concussion [97]. According to Navarro et al. [97], the year-over-year change in contract value for the concussion group resulted in a mean overall salary reduction of USD $300 k ± $1,300 k/year. Moreover, the performance score reduction for all offensive scoring players sustaining concussions was statistically significant (i.e., pre, 9 ± 6 fantasy football point "FFP"s/game; post, 7 ± 4 FFPs/game). These aforementioned factors supported the search for AI techniques and methods to assist in predicting the risk of a concussion in a more accurate way [74, 75]. In addition, AI techniques or methods on "technical and tactical analysis" area have been applied in American football players [46, 54]. This sample was composed by collegiate [54, 75], professional, [46] and retired professional [74]. The most frequently used methods in that sport were the support vector machine (50%) [74, 75] and decision tree classifier (25%) [46] with Markov process (25%) [54].

With regard to the rest of team sports, AI applications were reported in professional ice hockey (Markov process [56] and support vector machine [59, 60]), cricket (*K*-means clustering [49]), field hockey (decision tree classifier [47]), and rugby (artificial neural network [34]). All these AI applications were on the "technical and tactical analysis" area, confirming the importance of these aspects for success in the team sports [98]. With regard to injury risk prediction, the decision tree classifier has been used in professional rugby players using the "training load" as the predictor [72]. In beach volleyball players, the monitoring data from wearable sensors has been used to model the injury risk via artificial neural network [68]. These studies highlight the importance of adequate management of training load in these sports too [99–101] along with the use of technology and innovative approaches in data management in order to protect the athletes' health [102, 103].

Limitations

This article presents the AI techniques or methods mostly used to predict injury risk and sporting performance in team sport athletes from research published in peer-reviewed journals in the last 5 years. Whether the same techniques and methods can be applied in individual sports remains an unanswered question. In the manuscripts were found differences on sample sizes, where some samples were not as large as others. However, this may have been a consequence of the large number of elite athletes who were part of the pooled sample, where 76% of them were professional athletes and for some sports it is not as common to obtain large samples of professional athletes as others.

Conclusions

Our analysis showed that the AI techniques or methods for predicting injury risk and sporting performance mostly used in team sports were artificial neural

networks, decision tree classifier, support vector machine, and Markov process. The team sports with the most AI applications were soccer, basketball, handball, and volleyball. The current state of development in the area proposes a promising future with regard to AI use in team sports. Further evaluation research based on prospective methods is warranted to establish the predictive performance of specific AI techniques and methods.

Additional Files

Additional file 1: Table S1. Risk of bias score. (PDF 545 kb)

Additional file 2: Table S2. Summary of the AI studies. (PDF 662 kb)

Abbreviations
AI: Artificial intelligence; FFP: Fantasy football point; NCAA: National Collegiate Athletic Association; PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses; R&D: Research and Development

Authors' Contributions
JGC contributed to the design, collection, analysis, interpretation of data, and writing up of the paper. DOC contributed to the design, collection, analysis, interpretation of data, and writing up of the paper. TVS contributed to the design, collection, analysis, interpretation of data, and writing up of the paper. JCS contributed to the interpretation of data and writing up of the paper. ACMP contributed to the design, collection, analysis, interpretation of data, and writing up of the paper. GPN contributed to the design, interpretation of data, and writing up of the paper. All authors read and approved the final manuscript.

Availability of Data and Materials
After publication, all data necessary to understand and assess the conclusions of the manuscript are available to any reader of Sports Medicine-Open.

Ethics Approval and Consent to Participate
Not applicable.

Consent for Publication
All authors consent to the publication of the manuscript.

Competing Interests
The authors, João Gustavo Claudino, Daniel de Oliveira Capanema, Thiago Vieira de Souza, Julio Cerca Serrão, Adriano C. Machado Pereira, and George P. Nassis, declare that they have no competing interests.

Author details
[1]University of São Paulo, School of Physical Education and Sport - Laboratory of Biomechanics, Av. Prof. Mello de Morais, 65 – Cidade Universitária, São Paulo, São Paulo 05508-030, Brazil. [2]Research and Development Department, LOAD CONTROL, Contagem, Minas Gerais, Brazil. [3]Computing Department, Federal Center for Technological Education of Minas Gerais, Belo Horizonte, Brazil. [4]Computer Science Department, Federal University of Minas Gerais, Belo Horizonte, Brazil. [5]Department of Sports Science, City Unity College, Athens, Greece. [6]School of Physical Education & Sport Training, Shanghai University of Sport, Qingyuanhuan Rd 650, Yangpu District, Shanghai 200438, China.

References
1. Russell S, Norvig P. Artificial Intelligence: a modern approach. 3rd ed. New Jersey: Pearson; 2016.
2. Witten IH, Frank E, Hall MA, et al. Data Mining: practical Machine Learning tools and techniques. 4th ed. San Francisco: Morgan Kaufmann; 2017.
3. Zaki MJ, Meira Jr, W. Data Mining and analysis: fundamental concepts and algorithms. 1st ed. Cambridge: Cambridge University Press; 2014.
4. Passfield L, Hopker JG. A mine of information: can sports analytics provide wisdom from your data? Int J Sports Physiol Perform. 2017;12(7):851–5.
5. Rein R, Memmert D. Big data and tactical analysis in elite soccer: future challenges and opportunities for sports science. Springerplus. 2016;5(1):1410.
6. Blei DM, Smyth P. Science and data science. Proc Natl Acad Sci USA. 2017;201702076.
7. Bunker RP, Thabtah F. A machine learning framework for sport result prediction. Appl Comput Info. 2019;15(1):27–33.
8. Israni ST, Verghese A. Humanizing artificial intelligence. JAMA. 2019; 321(1):29–30.
9. Shortliffe EH, Sepúlveda MJ. Clinical decision support in the era of artificial intelligence. JAMA. 2018;320(21):2199–200.
10. Baker M. Data science: industry allure. Nature. 2015;520(7546):253–5.
11. Jordan MI1, Mitchell TM2. Machine learning: trends, perspectives, and prospects. Science. 2015;349(6245):255-60.
12. McCall A, Davison M, Carling C, et al. Can off-field 'brains' provide a competitive advantage in professional football? Br J Sports Med. 2016; 50(12):710–2.
13. Nassis GP. Leadership in science and medicine: can you see the gap? Sci Med Football. 2017;1(3):195–6.
14. Borresen J, Lambert MI. The quantification of training load, the training response and the effect on performance. Sports Med. 2009;39(9):779–95.
15. López-Valenciano A, Ayala F, Puerta JM, et al. A preventive model for muscle injuries: a novel approach based on learning algorithms. Med Sci Sports Exerc. 2018;50(5):915–27.
16. Novatchkov H, Baca A. Artificial intelligence in sports on the example of weight training. J Sports Sci Med. 2013;12(1):27–37.
17. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med. 2009;6(7):e1000097.
18. Saw AE, Main LC, Gastin PB. Monitoring the athlete training response: subjective self-reported measures trump commonly used objective measures: a systematic review. Br J Sports Med. 2016;50(5):281–91.
19. Horsley T, Dingwall O, Sampson M. Checking reference lists to find additional studies for systematic reviews. Cochrane Database Syst Rev. 2011; 10(8):MR000026.
20. Li C. Predict the neural network mathematical model of basketball team scores based on improved BP algorithm. BioTechnol Indian J. 2013;8(5):628–33.
21. Lu G. Evaluation model of young basketball players' physical quality and basic technique based on RBF neural network. BioTechnol Indian J. 2013;8(9):1193–8.
22. Wu L. The participating team's technical analysis of women's basketball in the 30th Olympic Games based on neural network. J Chem Pharma Res. 2013;5(11):152–8.
23. Zhang Q. Prediction based on basketball competition video athlete behaviors and on-line RBF neural network application. BioTechnol Indian J. 2014;10(4):764–9.
24. Kempe M, Grunz A, Memmert D. Detecting tactical patterns in basketball: comparison of merge self-organising maps and dynamic controlled neural networks. Eur J Sport Sci. 2015;15(4):249–55.
25. Bianchi F, Facchinetti T, Zuccolotto P. Role revolution: towards a new meaning of positions in basketball. Electron J Appl Stat Anal. 2017;10(3):712–34.
26. Tilp M, Schrapf N. Analysis of tactical defensive behavior in team handball by means of artificial neural networks. IFAC-PapersOnLine. 2015;28(1):784–5.
27. Hassan A, Schrapf N, Ramadan W, et al. Evaluation of tactical training in team handball by means of artificial neural networks. J Sports Sci. 2017;35(7):642–7.
28. Hassan A, Schrapf N, Tilp M. The prediction of action positions in team handball by non-linear hybrid neural networks. Int J Perform Anal Sport. 2017;17(3):293–302.

Claudino *et al. Sports Medicine - Open*   (2019) 5:28

Page 11 of 12

29. Schrapf N, Alsaied S, Tilp M. Tactical interaction of offensive and defensive teams in team handball analysed by artificial neural networks. Math Comput Model Dyn Syst. 2017;23(4):363–71.

30. Abdullah MR, Maliki ABHM, Musa RM, et al. Intelligent prediction of soccer technical skill on youth soccer player's relative performance using multivariate analysis and artificial neural network techniques. Int J Adv Sci Eng Inf Technol. 2016;6(5):668–74.

31. Strnad D, Nerat A, Kohek Š. Neural network models for group behavior prediction: a case of soccer match attendance. Neural Comput App. 2017;28(2):287–300.

32. Park YJ, Kim HS, Kim D, et al. A deep learning-based sports player evaluation model based on game statistics and news articles. Knowledge Syst. 2017;138:15–26.

33. Tümer AE, Koçer S. Prediction of team league's rankings in volleyball by artificial neural network method. Int J Perform Anal Sport. 2017;17(3):202–11.

34. Croft H, Lamb P, Middlemas S. The application of self-organising maps to performance analysis data in rugby union. Int J Perform Anal Sport. 2015;15(3):1037–46.

35. Fuster-Parra P, Garcia-Mas A, Cantallops J, et al. Ranking features on psychological dynamics of cooperative team work through Bayesian networks. Symmetry. 2016;8(5):34.

36. Link D, Hoernig M. Individual ball possession in soccer. PLoS One. 2017; 12(7):e0179953.

37. Healey G. Learning, visualizing, and assessing a model for the intrinsic value of a batted ball. IEEE Access. 2017;5:13811–22.

38. Jelinek HF, Kelarev A, Robinson DJ, et al. Using meta-regression data mining to improve predictions of performance based on heart rate dynamics for Australian football. Appl Soft Comput J. 2014;14:81–7.

39. Robertson S, Back N, Bartlett JD. Explaining match outcome in elite Australian Rules football using team performance indicators. J Sports Sci. 2016;34(7):637–44.

40. Woods CT, Veale J, Fransen J, et al. Classification of playing position in elite junior Australian football using technical skill indicators. J Sports Sci. 2018;36(1):97–103.

41. Leicht AS, Gomez MA, Woods CT. Explaining match outcome during the men's basketball tournament at the Olympic Games. J Sports Sci Med. 2017;16(4):468–73.

42. Leicht AS, Gomez MA, Woods CT. Team performance indicators explain outcome during women's basketball matches at the Olympic Games. Sports. 2017;5(4):e96.

43. Cene E. What is the difference between a winning and a losing team: insights from Euroleague basketball. Int J Perform Anal Sport. 2018; 18(1):55–68.

44. Montoliu R, Martin-Felez R, Torres-Sospedra J, et al. Team activity recognition in Association Football using a Bag-of-Words-based method. Hum Mov Sci. 2015;41:165–78.

45. Carpita M, Sandri M, Simonetto A, et al. Discovering the drivers of football match outcomes with data mining. Qual Technol Quant Manag. 2016;12(4):561–77.

46. Bock JR. Empirical prediction of turnovers in NFL Football. Sports. 2017;5(1):e1.

47. Morgan S, Williams MD, Barnes C. Applying decision tree induction for identification of important attributes in one-versus-one player interactions: a hockey exemplar. J Sports Sci. 2013;31(10):1031–7.

48. Hoch T, Tan X, Leser R, et al. A knowledge discovery framework for the assessment of tactical behaviour in soccer based on spatiotemporal data. Math Comput Model Dyn Syst. 2017;23(4):384–98.

49. Sankaran S. Comparing pay versus performance of IPL Bowlers: an application of cluster analysis. Int J Perform Anal Sport. 2014;14(1):174–87.

50. Wang M. Evaluating technical and tactical abilities of football teams in euro 2012 based on improved information entropy model and SOM neural networks. Int J Multimedia Ubiquitous Eng. 2014;9(11):293–302.

51. Vales-Alonso J, Chaves-Dieguez D, Pablo LM, et al. SAETA: a smart coaching assistant for professional volleyball training. IEEE Trans Syst Man Cybernet. 2015;45(8):1138–50.

52. Sheng L. Study of application of factors of volleyball game based on data mining. Info Technol J. 2013;12(19):5172–6.

53. Haiyan J, Jianqiang F. Application of data mining in technical and tactical analysis of volleyball match. Tech Bulletin. 2017;55(20):327–32.

54. Kolbush J, Sokol J. A logistic regression/Markov chain model for American college football. Int J Comput Sci Sport. 2017;16(3):185–96.

55. Zi M, Gao D. Technical and tactical analysis of basketball match based on data mining. Tech Bulletin. 2017;55(16):518–23.

56. Schulte O, Khademi M, Gholami S, et al. A Markov Game model for valuing actions, locations, and team performance in ice hockey. Data Min Knowl Discov. 2017;31(6):1735–57.

57. Bock JR. Pitch sequence complexity and long-term pitcher performance. Sports. 2015;3(1):40–55.

58. Soto VC. Predicting win-loss outcomes in MLB regular season games-a comparative study using data mining methods. Int J Comput Sci Sport. 2016;15(2):91–112.

59. Demers S. Riding a probabilistic support vector machine to the Stanley Cup. J Quant Anal Sports. 2015;11(4):205–18.

60. Gu W, Saaty TL, Whitaker R. Expert system for ice hockey game prediction: data mining with human judgment. Int J Inf Technol Dec Mak. 2016;15(4):763–89.

61. Wang Y, Zhao Y, Chan RHM, et al. Volleyball skill assessment using a single wearable micro inertial measurement unit at wrist. IEEE Access. 2018;6:13758–65.

62. Pai PF, ChangLiao LH, Lin KP. Analyzing basketball games by a support vector machines with decision tree model. Neural Comput Applicat. 2017; 28(12):4159–67.

63. Qilin S, Xiaomei W, Xiaoling F, et al. Study on knee joint injury in college football training based on artificial neural network. Rev Ibe Sist Tecnol Inf. 2016;E10:197–210.

64. Adetiba E, Iweanya VC, Popoola SI, et al. Automated detection of heart defects in athletes based on electrocardiography and artificial neural network. Cogent Eng. 2017;4(1):1411220.

65. Ertelt T, Solomonovs I, Gronwald T. Enhancement of force patterns classification based on Gaussian distributions. J Biomech. 2018;67:144–9.

66. Ge S. Research on the factors of basketball injury in physical teaching based on artificial neural network. Rev Facultad Ingen. 2017;32(3):415–22.

67. Bartlett JD, O'Connor F, Pitchford N, et al. Relationships between internal and external training load in team-sport athletes: evidence for an individualized approach. Int J Sports Physiol Perform. 2017;12(2):230–4.

68. Kautz T, Groh BH, Hannink J, et al. Activity recognition in beach volleyball using a deep convolutional neural network. Data Min Knowl Discov. 2017;31(6):1678–05.

69. Pensgaard AM, Ivarsson A, Nilstad A, et al. Psychosocial stress factors, including the relationship with the coach, and their influence on acute and overuse injury risk in elite female football players. BMJ Open Sport Exerc Med. 2018;4(1):e000317.

70. Ruddy JD, Shield AJ, Maniar N, et al. Predictive modeling of hamstring strain injuries in elite Australian Footballers. Med Sci Sports Exerc. 2018;50(5):906–14.

71. Rossi A, Pappalardo L, Cintia P, et al. Effective injury forecasting in soccer with GPS training data and machine learning. PLoS One. 2018;13(7):e0201264.

72. Thornton HR, Delaney JA, Duthie GM, et al. Importance of various training-load measures in injury incidence of professional Rugby League athletes. Int J Sports Physiol Perform. 2017;12(6):819–24.

73. Jaspers A, De Beéck TO, 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):625–30.

74. Goswami R, Dufort P, Tartaglia MC, et al. Frontotemporal correlates of impulsivity and machine learning in retired professional athletes with a history of multiple concussions. Brain Struct Funct. 2016;221(4):1911–25.

75. Wu LC, Kuo C, Loza J, et al. Detection of American football head impacts using biomechanical features and support vector machine classification. Sci Rep. 2018;8(1):855.

76. Whiteside D, Martini DN, Lepley AS, et al. Predictors of ulnar collateral ligament reconstruction in Major League Baseball pitchers. Am J Sports Med. 2016;44(9):2202–9.

77. Lapham AC, Bartlett RM. The use of artificial intelligence in the analysis of sports performance: a review of applications in human gait analysis and future directions for sports biomechanics. J Sports Sci. 1995;13(3):229–37.

78. Zelic I, Kononenko I, Lavrac N, et al. Induction of decision trees and Bayesian classification applied to diagnosis of sport injuries. J Med Syst. 1997;21(6):429–44.

79. Mikołajec K, Maszczyk A, Zając T. Game indicators determining sports performance in the NBA. J Hum Kinet. 2013;37:145–51.

80. Leite N, Coelho E, Sampaio J. Assessing the importance given by basketball coaches to training contents. J Hum Kinet. 2011;30:123–33.

81. Przednowek K, Krzeszowski T, Przednowek KH, Lenik P. A system for analysing the basketball free throw trajectory based on particle swarm optimization. Appl Sci. 2018;8(11):2090.

82. de Loës M, Dahlstedt LJ, Thomée R. A 7-year study on risks and costs of knee injuries in male and female youth participants in 12 sports. Scand J Med Sci Sports. 2000;10(2):90–7.

83. Tirabassi J, Brou L, Khodaee M, et al. Epidemiology of high school sports-related injuries resulting in medical disqualification: 2005-2006 through 2013-2014 academic years. Am J Sports Med. 2016;44(11):2925–32.

84. Harmon KG, Asif IM, Maleszewski JJ, et al. Incidence, cause, and comparative frequency of sudden cardiac death in National Collegiate Athletic Association athletes: a decade in review. Circulation. 2015; 132(1):10–9.

85. Wasfy MM, Hutter AM, Weiner RB. Sudden cardiac death in athletes. Methodist Debakey Cardiovasc J. 2016;12(2):76–80.

86. Malfait B, Dingenen B, Smeets A, et al. Knee and hip joint kinematics predict quadriceps and hamstrings neuromuscular activation patterns in drop jump landings. PLoS One. 2016;11(4):e0153737.

87. Taylor JB, Ford KR, Nguyen AD, et al. Biomechanical comparison of single- and double-leg jump landings in the sagittal and frontal plane. Orthop J Sports Med. 2016;4(6):2325967116655158.

88. Bahr R. Why screening tests to predict injury do not work-and probably never will…: a critical review. Br J Sports Med. 2016;50(13):776–80.

89. Gaudet S, Begon M, Tremblay J. Cluster analysis using physical performance and self-report measures to identify shoulder injury in overhead female athletes. J Sci Med Sport. 2018;S1440-2440(18):30863–6.

90. Afonso J, Esteves F, Araújo R, et al. Tactical determinants of setting zone in elite men's volleyball. J Sports Sci Med. 2012;11(1):64–70.

91. Wagner H, Finkenzeller T, Würth S, et al. Individual and team performance in team-handball: a review. J Sports Sci Med. 2014;13(4):808–16.

92. Carey DL, Crow J, Ong KL, et al. Optimizing preseason training loads in Australian Football. Int J Sports Physiol Perform. 2018;13(2):194–9.

93. Fortington LV, Berry J, Buttifant D, et al. Shorter time to first injury in first year professional football players: a cross-club comparison in the Australian Football League. J Sci Med Sport. 2016;19(1):18–23.

94. Stares J, Dawson B, Peeling P, et al. Identifying high risk loading conditions for in-season injury in elite Australian football players. J Sci Med Sport. 2018;21(1):46–51.

95. Whiteley R. 'Moneyball' and time to be honest about preseason screening: it is a sham making no inroads on the 1 billion dollar injury costs in baseball. Br J Sports Med. 2016;50(14):835–6.

96. Reider B. Moneyball. Am J Sports Med. 2014;42(3):533–5.

97. Navarro SM, Sokunbi OF, Haeberle HS, et al. Short-term outcomes following concussion in the NFL: a study of player longevity, performance, and financial loss. Orthop J Sports Med. 2017;5(11): 2325967117740847.

98. Halouani J, Chtourou H, Gabbett T, et al. Small-sided games in team sports training: a brief review. J Strength Cond Res. 2014;28(12):3594–618.

99. Dalton-Barron NE, McLaren SJ, Black CJ, et al. Identifying contextual influences on training load: an example in professional Rugby Union. J Strength Cond Res. 2018. https://doi.org/10.1519/JSC.0000000000002706.

100. McLaren SJ, Weston M, Smith A, et al. Variability of physical performance and player match loads in professional rugby union. J Sci Med Sport. 2016;19(6):493–7.

101. Oliveira WK, Jesus K, Andrade AD, et al. Monitoring training load in beach volleyball players: a case study with an Olympic team. Motriz. 2018;24(1):e1018155.

102. Düking P, Achtzehn S, Holmberg HC, Sperlich B. Integrated framework of load monitoring by a combination of smartphone applications, wearables and point-of-care testing provides feedback that allows individual responsive adjustments to activities of daily living. *Sensors (Basel)*. 2018;18(5):E1632.

103. Austen K. What could derail the wearables revolution? Nature. 2015; 525(7567):22–4.

Publisher's Note