Artificial Intelligence for Sport Injury Prediction

Chapter · April 2024

DOI: 10.1007/978-3-031-67256-9_5

3 authors:

Robin Owen
Liverpool Hope University

Julian Andrew Owen
Bangor University

Seren Evans
Bangor University

Book chapter title:

Artificial Intelligence for Sport Injury Prediction

Book chapter authors:

Robin Owen[1], Julian A Owen[2], and Seren L Evans[2]

1. Liverpool Hope University, UK
2. Bangor University, UK

From the book:

Artificial Intelligence in Sports, Movement, and Health

Book editors:

Carlo Dindorf, Eva Bartaguiz, Freya Gassmann, and Michael Fröhlich

Publisher:

Springer

Keywords:

Machine Learning, Pattern Recognition, Sport Injury, Injury Prevention

Abstract:

Preventing injury is a core facilitator of success in sport. Thus, vast sums of money are invested into achieving this. However, sport injury is still seen as equal parts ‘art’ and science. Despite the best efforts of individuals, teams, and national bodies to apply scientifically-derived injury prevention strategies, millions of athletes still get injured in sport every year. Evidently, sport injury prediction is a field which has scope for improvement. One potential way of advancing the field is the use of AI (artificial intelligence). It offers an opportunity to: (1) treat sporting injury as the complex phenomenon it appears to be; (2) consider the non-linear context surrounding athlete injuries; and (3) provide a supplement to practitioner reasoning, to facilitate quicker decisions. The present book chapter evaluates previous research studies’ use of AI for injury prediction, assesses the unique advantages offered by AI-based analyses, and discusses challenges when attempting to utilise AI for injury prediction. Overall, the use of AI for sport injury prediction offers a fascinating opportunity. It may one day create a revolution in the field, improving not only prediction itself but also our understanding of the complex interactive factors which govern injury in sport.

Sport Injury - The Context

It is well established that participation in sports offers numerous physical and mental health benefits alongside providing opportunities for social interaction and the development of positive psychosocial health (Eime et al., 2013). However, the benefits of sport participation are accompanied by a significant sport-related injury burden in both elite and recreational athletes (Emery et al., 2007; Jacobsson et al., 2012). Despite this, there is a relative paucity of research evaluating the efficacy of injury prevention strategies (Conn et al., 2003).

The exact number of sports injuries worldwide each year is challenging to determine precisely due to variations in reporting systems, definitions of sports injuries, and the vast range of sports and activities involved. Estimates suggest that sport-related injuries are common with millions of people, suffering from injuries each year, ranging from minor sprains and strains to more severe fractures, concussions, and other traumatic injuries. In context, an estimated seven million Americans and almost six million Europeans receive medical attention annually for sport-related injuries (Conn et al., 2003; Kisser and Bauer, 2012). Roughly one in five school children miss at least one day of school, while one in three working adults loses at least one workday yearly due to sport-related injuries (Conn et al., 2003; Emery et al., 2006).

Advancements in comprehending the financial strain and allocating resources toward preventing sports injuries have been constrained, partly due to difficulties in clearly defining the extent, breadth, and financial implications of the sports injury issue. An Australian research study approximated the burden of sports-related injuries over a span of seven years to amount to $265 million Australian dollars (Finch et al., 2015). In Europe, the economic assessment of health expenditures, considering both the savings generated through sports participation and the losses incurred due to injuries, suggests that 40-50% of the economic advantage is eroded by sports-related injuries (Weiß, 2000; BASBO, 2001). Many of these estimates of the direct costs represent medical related treatment costs and ignore the indirect costs which include the immediate and future loss of income costs due to injury. Therefore, the financial cost of sport-injury is likely underestimated as indirect costs can account for approximately 46–71 % of the total costs associated with injuries (Lacny et al., 2014).

The repercussions of sports injuries extend beyond mere physical and financial implications. It is widely acknowledged that there exists a significant emotional and psychological toll on athletes' mental health and well-being. This toll often manifests in the form of depression, stress, anger, and diminished self-esteem, especially among competitive athletes or those severely injured (Smith, 1996). Therefore, as sport and physical activity continues to be promoted as part of a healthy lifestyle, sport-related injuries are becoming an important public health concern.

In competitive sports, the adverse effects of injuries are typically more pronounced. It is recognised that the burden of injuries escalates with the level of competition, primarily due to greater exposure to rigorous training and competitions, leading to increased physical and psychological strain. Professional and national sports organisations are obligated to ensure the well-being of their athletes; hence, prioritising athlete welfare is crucial. Lowering the burden of injuries also becomes a notable advantage for team success, which influences commercial revenues.

Injury prediction should be a key component for injury prevention, since the successful identification of injury predictors forms the basis for effective preventive measures. Traditionally, research focusing on the prevention of sports injury is based on the ‘sequence of prevention’ which includes injury audit (surveillance) to establish the extent and nature of the problem, identification of risk factors, and implementation of relevant prevention strategies based on these findings (van Mechelen, 1992). This epidemiological approach is useful as it allows researchers to identify the risk of injury (injury incidence or rate and injury burden), prevalence, and risk factors associated with injury within different sports and populations and helps to identify patterns and trends, contributing to injury occurrence. This approach has often attested that single risk factors account for the occurrence of an injury. Although this approach has uncovered numerous potential predictors of injuries using conventional statistical methods like logistic regression. Unsurprisingly, these methods have not consistently identified risk factors (Bekker and Clark., 2016). These inconsistencies underscore the complexity inherent in most human health conditions.

Fundamentally, sports injury is a multifaceted phenomenon influenced by various modifiable and non-modifiable risk factors, including biomechanical, physiological, psychological, environmental, and sociocultural aspects. To understand injury risk, we must analyse the forces, loads, and motions involved in sports activities to understand how they contribute to tissue stress, strain, and injury. We must consider the psychological factors that can modulate the physiological responses to stress and influence injury vulnerability. We must also include context and consider the influence of societal values, gender roles, coach-athlete relationships, peer interactions, and institutional practices on athlete behaviour, risk-taking, and injury reporting.

Since the ‘sequence of prevention’ was first suggested, several models have been developed to conceptualise the complexities surrounding sports injury occurrence and that the injury has a non-linear behaviour (Meeuwisse., 1994; Meeuwisse et al., 2007; Bittencourt et al., 2016; Bekker and Clark., 2016). These models suggest that the multifaceted and intricate nature of sports injuries does not stem solely from the linear combination of isolated predictive factors but rather from the interplay often referred to as 'the web of determinants' (Philippe and Mansi, 1998). These determinants may be interconnected in a nonlinear fashion, meaning that slight changes in a few determinants can result in significant and occasionally unforeseen outcomes. To comprehensively understand the complex origins of sports injuries, a complex systems approach is essential.

The Current State of AI for Injury Prediction

As outlined, it is well established that sports injuries are multifactorial in nature, and very rarely are attributed to a singular variable in the line of causation; rather, sports injuries arise from multiple interactions between both modifiable (i.e. training load, strength) and non-modifiable determinants (i.e. age, previous injury history) and their non-linear fluctuations over time (Bittencourt et al., 2016; Hulme & Finch, 2015). Therefore, to accurately determine the complexity of their origin, sports injury prediction requires a complex systems approach to better understand how these intricate interactions lead to injury.

Recent advancements of artificial intelligence (AI) based analysis (including machine learning and pattern recognition) have lead to its introduction into the realm of sports medicine research (Ruddy et al., 2018; Van Eetvelde., 2021), allowing for a more robust analysis of large quantities of data to formulate prediction models of injury (Sigurdson & Chan, 2021). AI can be designed to process imbalanced datasets, which is commonplace in sports injury research as, typically there will be more athletes not sustaining an injury when compared to those sustaining an injury (Lopez-Valenciano et al., 2018; Van Eetvelde et al., 2021). Furthermore, utilising AI for sports injury research allows for the inclusion of both modifiable and non-modifiable risk factors as input features and can be used to evaluate their effectiveness in predicting injury as a binary classification outcome (injury versus no injury). Caution is needed that we are not reverting back to over-simplistic, reductionist views of injuries, such as injuries occuring due to singular inciting events. Models which have previously been generated for targeted injury diagnoses (for example; lower extremity injuries, lateral ankle sprains) may be of greater sensitivity within multivariate modelling when compared to grouping all injuries together, producing more interpretable and unambiguous findings for injury incidence (Henriquez et al., 2020).

Various predictive variables of sports injury have, therefore, emerged across a range of sports as a result of AI-based analyses. Within Australian Football, risk factors such as age, stature, body mass, playing position, and previous lower limb injury history were identified as predictors of hamstring strain injury, with an associated accuracy of 85% across algorithms (Ruddy et al., 2018); namely, Naive Bayes, Logistic Regression, Random Forest, Support Vector Machine, and Neural Network, which have qualities of probabilistic classification and the ability to model complex, non-linear interactions within multiple predictor variables (John & Langley, 1994; Quinlan, 1993; Keerthi et al., 2006). Utilising a similar approach with random forest algorithms in identifying lower limb musculoskeletal injuries amongst NCAA athletes, Rommers et al. (2020) identified hip-based strength metrics, demographic and balance variables as indicators for future injury. Furthermore, adopting a subgroup discovery approach which allows for the analysis of subsets of individuals who share common attributes for injury risk from input features (Herrera et al., 2011), de Leeuw et al. (2022) discovered that predictors of injury within elite male volleyball were fatigue, overuse, sleep, muscle soreness, and training exertion. Physical attributes such as height and weight, alongside strength, flexibility, speed, agility, and endurance features, achieved 85% precision using XGBoost in assessing injury predictors within elite youth football (Rommers et al., 2020). Pattern recognition analyses, therefore, show initial potential to provide a feasible statistical method of forecasting injuries in sport whilst being able to account for (1) modifiable and non-modifiable risk factors, (2) the time-series nature of athlete training data, (3) whilst also considering their nonlinear interactions.

Advantages of using AI for Injury Prediction

Natural sciences, such as injury prediction in sport, traditionally adhere to explanatory positivist views where understanding and generalisation of phenomena require the testing of clearly defined hypotheses (i.e., predictions) using tightly controlled methods (Kuhn, 1962). This approach inherently encourages a ‘reductionist’ approach to research, wherein testing theoretically-based and limited-in-number predictors of phenomena is considered superior. Indeed, injury prediction research has predominantly adhered to these principles (Bekker & Clark, 2016; Bittencourt et al., 2016). However, such approaches induce a case of ‘survivor bias’; factors are prioritised for consideration if their relationship with injury is either known or can be clearly predicted (Lockwood, 2021). Consequently, this may prohibit the identification of new, as-of-yet unknown, factors which may affect injury in sport (Tee et al., 2020). Similarly, once a certain number of predictors is reached, it can make it difficult for researchers to fully grasp their interaction.

Concerted efforts to broaden understanding of sports injury are of particular importance given recent calls to consider sports injury as a complex phenomenon, affected by many variables and interactions (Fonseca et al., 2020; Tee et al., 2020). Explanatory positivist approaches to-date have laid the foundations for identifying modifiable and non-modifiable risk factors of injury in sport (Bahr, 2016; Rossi et al., 2021), but limitations arise from the typical utilisation of mono-dimensional approaches. Variables are often treated as static, absolute at one point in time, and subsequently ignore the complex underlying pattern of sports injuries and time-series nature of athlete status (Rossi et al., 2021). This ‘static’ attitude to predictors, combined with assumptions of linear relationships between singular variables and injury, means that current approaches with high explanatory power do not always translate to high predictive power in relation to injury risk (Jauhianen et al., 2021; Shmueli, 2010).

Therefore, a possible means to deepening understanding of injury predictors are AI-assisted analyses. AI is particularly suited to complex problems, given its ability to: process large volumes of data; comprise partial automation to reduce time cost; provide non-linear assessment of multiple interactions; and discover useful hidden patterns in data (Zhuang et al., 2017; Pham et al., 2020). Accordingly, sports injury researchers are beginning to utilise artificial neural networks, support vector machines, gradient boosting machines, and decision tree methods (Bullock et al., 2022). Although pursuing complex analytical procedures such as these goes against fundamental scientific principles (e.g., Occam’s Razor, wherein the simplest explanations are regarded as the most plausible, and should thus be pursued; Blumer et al., 1987), injury risk appears to be highly complex by nature (Fonseca et al., 2020; Tee et al., 2020) and may thus benefit from AI-assisted analyses. Specifically, AI-based approaches could demonstrate a better capacity for interpreting the highly complex and non-linear contexts surrounding each case despite their seeming contradiction with established explanatory conventions (Tee et al., 2020).

Training and Testing an AI for Injury Prediction

Just like the athletes themselves, AI models require rigorous training and testing (Kanal & Chandrasekaran, 1971). ‘Training’ entails calibrating the underlying parameters which AI models use to produce outputs from inputs. ‘Testing’ entails evaluating the effectiveness of these models, often using a different dataset to that used in training. There are many testing/training methods, such as supervised learning, unsupervised learning, and reinforcement learning; however, a commonality among them all is a requirement for large volumes of representative data to create models which provide accurate outputs (L’heureux et al., 2017). Although the quantity and quality of training/testing data is only one of many factors which can cause undesirable bias in models, it is one of the key determinants (see Prediction model Risk Of Bias Assessment Tool; Wolff et al., 2019). If an AI model is subject to insufficient volumes of relevant data during training, it is likely that these models will contain bias, which can lead to inaccurate outputs.

A recent systematic review found that 98% of AI-based analyses used to predict sporting injuries were at high or unclear risk of bias (Bullock et al., 2022). In part, this is a product of the additional challenges the field has when it comes to testing and training models; contexts surrounding injury are dynamic and not interchangeable (Tee et al., 2020). Injury can be affected by more than just match play and training load. It is highly dependent on the context surrounding an athlete. Historical, political, social, economic, scientific, cultural, and organisational factors can all affect injury likelihood and the effectiveness of preventative methods. For example, playing on hard ground out of geographical/economic necessity can increase injury likelihood (Chalmers et al., 2012). Relatedly, the contexts surrounding injury are dynamic rather than static. For example, Between 1998 and 2010, rugby union forwards have become 22% heavier, 8% taller, and 18% stronger (Lombard et al., 2015). Likewise, changes in coaching and backroom staff can produce profound changes to an athlete’s recovery protocols from one year to the next (Galdino et al., 2023). A result of the complex and dynamic factors surrounding sports injury is that it makes it challenging for a single research team to collect sufficient predictors as well as sufficient volumes of data to optimally train and test AI models.

Given the challenges faced when attempting to apply AI methods to sports injury, it is not surprising that previous studies have been criticised for their generalisability and application to applied contexts (see Bullock et al., 2022). Specifically, it has been suggested that even in AI-based studies featuring low risk of bias, modest predictive performance of models means that there may be no injury prediction models which can be confidently recommended for applied practice. Going forward, it may be necessary for researchers to embrace the open science to collaborate and compile sufficiently detailed datasets. Such Open Science approaches entail intentional sharing of data (and failing that, making data freely accessible) to better build on previous studies (Vicente-Saez & Martinez-Fuentes, 2018). Precedent for the rapid advancement of AI given sufficient access to detailed datasets for testing/training can be seen in text-to-text applications such as Chat GPT, where access to a large corpus of written work throughout history has allowed impressively accurate predictions of desired text, based on user inputs (Wu et al., 2023). Researchers investigating sports injury should aim to work together to further elucidate underlying complex interactions of predictors.

Practical Implications: Pitfalls and Solutions

Although AI has the potential to become a powerful tool in injury prediction (Bullock et al., 2022), its underlying mechanisms may be too complex for applied practitioners to find useful/comprehensible themselves. Therefore, AI-based approaches may further increase the researcher-practitioner gap. This researcher-practitioner gap occurs when scientifically derived knowledge is not applied by practitioners in the field (Lenfant, 2003). The present wealth of different AI-based approaches, complex statistical metrics, and frequent requirement to modify computer code, means that a majority of applied practitioners may struggle to use AI models in any capacity other than standardised 'plug and play'' packages (Bullock et al., 2022). However, even if 'plug and play' packages are made available to applied practitioners, current sports injury models' high likelihood of bias (Bullock et al., 2022) run a high risk of incorrect application. In such high risk situations, it has been shown that individuals tend to rely on their own judgement and avoid applying these high risk methods, further widening the researcher-practitioner gap (Jøsang & Presti, 2004; Papenmeier et al., 2022). Therefore, in addition to producing accurate injury prediction models, another key barrier may be to overcome the researcher-practitioner gap.

The utilisation of AI-based analyses in injury prediction studies is often hindered by limited data inclusion, restricting analysis to a narrow scope of variables. For instance, some studies only incorporate physical performance metrics (Rommers et al., 2020), perhaps constraining predictive accuracy. However, the potential for heightened precision remains, suggesting an opportunity for enhancement through integrating more extensive datasets (Verhagen & Bolling, 2015). By refining the focus of injury prediction using advanced AI methodologies, such as targeting specific injury types prevalent within distinct athletic cohorts - such as hamstring strains in elite football or anterior cruciate ligament injuries in female athletes - the applicability of these models to real-world practice can be improved (Van Eetvelde et al., 2021; Rommers et al., 2020). This may provide practitioners with more robust datasets, enabling the implementation of more effective and targeted injury prevention strategies.

That said, to create more accurate prediction models, reduce bias, promote practitioner uptake, and reduce the researcher-practitioner gap, theoretically driven variables of injury risk factors still require prioritisation when deciding on input features during preprocessing stages of AI analyses. To illustrate, a strong relationship exists between the amount of ice cream sold and shark attack incidences, and it may even be possible to predict the number of shark attacks that will occur based on the number of ice creams sold. However, in reality, no amount of regulating ice cream sales will have an effect on the number of shark attacks; ice cream sales are epiphenomenal to shark attacks and is likely a byproduct of another process, such as warmer weather resulting in more demand for ice cream and people visiting the beach. Regulating the waters with more coastal surveillance and warnings for surfers is likely to be more effective in reducing the number of shark attacks. The point is, utilising variables that are theoretically linked to sports injury will reduce the likelihood of erroneous discoveries, which would affect the interpretability and reliability of the models.

Conclusion

AI-based approaches to sports injury prediction provide many opportunities to advance the field. Firstly, it has the capacity to treat sporting injury as the complex phenomenon it appears to be. Secondly, it allows for consideration of the non-linear context surrounding athlete injuries, which previous reductionist statistical approaches were forced to omit. Lastly, it can provide a supplement to practitioner reasoning, to facilitate quicker decisions. However, one should not overlook the challenges of using AI. Training effective AI requires large and representative datasets, which has been a key barrier faced in sports injury research. Additionally, until accurate models become available as 'plug and play' solutions, they may be prohibitively complex/novel for applied practitioners to use; thus potentially widening researcher-practitioner gaps. If these challenges are overcome though, AI may one day revolutionise not only sports injury prediction accuracy, but also our understanding of underlying factors and their interaction.

Prediction models may, therefore prompt early intervention and manipulation of variables which are known to have an effect on injury risk however unless the relationship is causal, manipulating certain metrics does not mean that injury risk will be altered (Hernan et al., 2019); therefore, assuming that manipulating certain variables reduces the risk of injury is the equivalent of banning ice cream sales to prevent shark attacks (Impellizzeri et al., 2020). When handling data regarding injury prediction and prevention, identifying the optimal set of risk factors for athletes at greater risk of injury would prove invaluable for coaches, medical practitioners, and for the overall well-being of athletes. Achieving this necessitates a tailored approach to athlete monitoring practices and addressing key performance indicators tailored to the demands of each individual sport. Within the realm of sports, the cost of injury - weighing the costs of medical procedures, rehabilitation, player time loss due to injury, and its impact of team success against the benefits of injury reduction - is pivotal in the decision making process (Gabbett et al., 2016). When utilising AI, more efforts need to be made in relation to understand the relative weight of individual risk factors and injury risk, portraying a picture of the probability of injury rather than classifying an athlete into a high or low risk group (Van Eetvelde et al., 2021; Rossi et al., 2018), which would be of more benefit for sporting practitioners when it comes to making adjustments to training regimes and team selection. Employing an AI approach to injury management should, therefore, not only be able to identify risk factors but also provide practitioners with actionable thresholds for heightened injury probability, allowing for the implementation of timely prevention strategies with the hope of minimising the cost of injury for both athlete and team.

References

BASPO - Bundesamt für Sport. (2001). Volkswirtschaftlicher Nutzen der Gesundheitseffekte der körperlichen Aktivität: erste Schätzungen für die Schweiz. *Schweizer Zeitschrift für Sportmedizin und Sporttraumatologie, 49*(2), 84-86.

Bekker, S., & Clark, A. M. (2016). Bringing complexity to sports injury prevention research: from simplification to explanation. *British Journal of Sports Medicine*, *50*(24), 1489-1490.

Blumer, A., Ehrenfeucht, A., Haussler, D., & Warmuth, M. K. (1987). Occam's razor. *Information Processing Letters, 24*(6), 377-380.

Bittencourt, N. F., Meeuwisse, W. H., Mendonça, L. D., Nettel-Aguirre, A., Ocarino, J. M., & Fonseca, S. T. (2016). Complex systems approach for sports injuries: moving from risk factor identification to injury pattern recognition—narrative review and new concept. *British Journal of Sports Medicine*, *50*(21), 1309-1314.

Bullock, G. S., Mylott, J., Hughes, T., Nicholson, K. F., Riley, R. D., & Collins, G. S. (2022). Just how confident can we be in predicting sports injuries? A systematic review of the methodological conduct and performance of existing musculoskeletal injury prediction models in sport. *Sports Medicine, 52*(10), 2469-2482.

Chalmers, D. J., Samaranayaka, A., Gulliver, P., & McNoe, B. (2012). Risk factors for injury in rugby union football in New Zealand: a cohort study. *British Journal of Sports Medicine, 46*(2), 95-102.

Conn, J. M., Annest, J. L., Gilchrist, J. (2003). Sports and recreation related injury episodes in the US population, 1997-99. *Injury Prevention, 9*(2), 117-23

de Leeuw, A. W., van der Zwaard, S., van Baar, R., & Knobbe, A. (2022). Personalized machine learning approach to injury monitoring in elite volleyball players. *European Journal of Sport Science, 22*(4), 511-520.

Emery, C. A., Rose, M. S., McAllister, J. R., et al. (2007). A prevention strategy to reduce the incidence of injury in high school basketball: A cluster randomized controlled trial. *Clinical Journal of Sport Medicine, 17*, 17-24.

Emery, Carolyn A BScPT, PhD; Meeuwisse, Willem H MD, PhD; McAllister, Jenelle R BKin. (2006). Survey of Sport Participation and Sport Injury in Calgary and Area High Schools. *Clinical Journal of Sport Medicine, 16*(1), 20-26.

Eime, R. M., Young, J. A., Harvey, J. T., et al. (2013). A systematic review of the psychological and social benefits of participation in sport for children and adolescents: Informing development of a conceptual model of health through sport. *International Journal of Behavioral Nutrition and Physical Activity, 10*, 98.

Finch, C. F., Kemp, J. L., Clapperton, A. J. (2015). The incidence and burden of hospital-treated sports-related injury in people aged 15+ years in Victoria, Australia, 2004-2010: A future epidemic of osteoarthritis? *Osteoarthritis Cartilage, 23*(7), 1138-43.

Fonseca, S. T., Souza, T. R., Verhagen, E., Van Emmerik, R., Bittencourt, N. F., Mendonça, L. D., ... & Ocarino, J. M. (2020). Sports injury forecasting and complexity: a synergetic approach. *Sports Medicine, 50*, 1757-1770.

Gabbett, H. T., Windt, J., & Gabbett, T. J. (2016). Cost-benefit analysis underlies training decisions in elite sport. *British Journal of Sports Medicine, 50*(21), 1291-1292.

Galdino de Souza, M., & Wicker, P. (2023). A culture of constraints: How head coach turnovers affect the backroom staff and player development in professional football. *International Journal of Sports Science and Coaching. 19*(2), 515-524.

Henriquez, M., Sumner, J., Faherty, M., Sell, T., & Bent, B. (2020). Machine learning to predict lower extremity musculoskeletal injury risk in student athletes. *Frontiers in Sports and Active Living, 2*, 576655.

Herrera, F., Carmona, C. J., González, P., & Del Jesus, M. J. (2011). An overview on subgroup discovery: foundations and applications. *Knowledge and Information Systems, 29*, 495-525.

Hernán, M. A., Hsu, J., & Healy, B. (2019). A second chance to get causal inference right: a classification of data science tasks. *Chance, 32*(1), 42-49.

Impellizzeri, F. M., Menaspà, P., Coutts, A. J., Kalkhoven, J., & Menaspà, M. J. (2020). Training load and its role in injury prevention, part I: back to the future. *Journal of Athletic Training, 55*(9), 885-892.

Jøsang, A., & Presti, S. L. (2004, March*). Analysing the relationship between risk and trust. In International conference on trust management* (pp. 135-145). Berlin, Heidelberg: Springer Berlin Heidelberg.

Jacobsson, J., Timpka, T., Kowalski, J., et al. (2012). Prevalence of musculoskeletal injuries in Swedish elite track and field athletes. *American Journal of Sports Medicine, 40*, 163-169.

Kanal, L., & Chandrasekaran, B. (1971). On dimensionality and sample size in statistical pattern classification. *Pattern Recognition, 3*(3), 225-234.

Keerthi, S. S., Chapelle, O., DeCoste, D., Bennett, K. P., & Parrado-Hernández, E. (2006). Building support vector machines with reduced classifier complexity. *Journal of Machine Learning Research, 7*(7).

Kisser, R., & Bauer, R. (2012). The burden of sports injuries in the European Union. *Austrian Road Safety Board,* 1-94.

Kuhn, T. S. (2012). *The structure of scientific revolutions*. University of Chicago press.

Lacny, S., Marshall, D. A., Currie, G., et al. (2014). Reality check: The cost–effectiveness of removing body checking from youth ice hockey. *British Journal of Sports Medicine, 48,* 1299-1305.

Lombard, W. P., Durandt, J. J., Masimla, H., Green, M., & Lambert, M. I. (2015). Changes in body size and physical characteristics of South African under-20 rugby union players over a 13-year period. *The Journal of Strength & Conditioning Research, 29*(4), 980-988.

López-Valenciano, A., Ayala, F., Puerta, J. M., Croix, M. D. S., Vera-García, F., Hernández-Sánchez, S., ... & Myer, G. (2018). A preventive model for muscle injuries: a novel approach based on learning algorithms. *Medicine and science in Sports and Exercise, 50*(5), 915.

L'heureux, A., Grolinger, K., Elyamany, H. F., & Capretz, M. A. (2017). Machine learning with big data: Challenges and approaches. *Ieee Access, 5,* 7776-7797.

Meeuwisse, W. (1994). Assessing causation in sport injury: A multifactorial model. Clinical *Journal of Sport Medicine, 4,* 66–170.

Meeuwisse, W. H., Tyreman, H., Hagel, B., et al. (2007). A dynamic model of etiology in sport injury: The recursive nature of risk and causation. *Clinical Journal of Sport Medicine, 17,* 215–19. do

Papenmeier, A., Kern, D., Englebienne, G., & Seifert, C. (2022). It's complicated: The relationship between user trust, model accuracy and explanations in ai. *ACM Transactions on Computer-Human Interaction (TOCHI), 29*(4), 1-33.

Pham, Q. V., Nguyen, D. C., Huynh-The, T., Hwang, W. J., & Pathirana, P. N. (2020). Artificial intelligence (AI) and big data for coronavirus (COVID-19) pandemic: a survey on the state-of-the-arts. *IEEE access, 8,* 130820-130839.

Philippe, P., & Mansi, O. (1998). Nonlinearity in the epidemiology of complex health and disease processes. *Theoretical Medicine and Bioethics, 19*(6), 591-607.

Quinlan, J. R. (1993). *Programs for machine learning*. Morgan Kauffmann Publishers.

Rommers, N., Rössler, R., Verhagen, E., Vandecasteele, F., Verstockt, S., Vaeyens, R., Lenoir, M., D'Hondt, E., & Witvrouw, E. (2020). A Machine Learning Approach to Assess Injury Risk in Elite Youth Football Players. *Medicine & Science in Sports & Exercise, 52*(8), 1745–1751.

Ruddy, J. D., Shield, A. J., Maniar, N., Williams, M. D., Duhig, S. J., Timmins, R. G., ... & Opar, D. A. (2018). Predictive modeling of hamstring strain injuries in elite Australian footballers. *Medicine & Science in Sports & Exercise, 50*(5), 906-914.

Sigurdson, H., & Chan, J. H. (2021). Machine Learning Applications to Sports Injury: A Review. *icSPORTS,* 157-168.

Smith, A. M. (1996). Psychological impact of injuries in athletes. *Sports Medicine, 22*, 391–405.

Tee, J. C., McLaren, S. J., & Jones, B. (2020). Sports injury prevention is complex: we need to invest in better processes, not singular solutions. *Sports Medicine*, *50*(4), 689-702.

Van Eetvelde, H., Mendonça, L. D., Ley, C., Seil, R., & Tischer, T. (2021). Machine learning methods in sport injury prediction and prevention: a systematic review. *Journal of experimental orthopaedics*, *8*, 1-15.

Van Mechelen, W., Hlobil, H., Kemper, H. (1992). Incidence, severity, etiology and prevention of sports injuries - A review of concepts. *Sports Medicine, 14*, 82–99.

Vicente-Saez, R., & Martinez-Fuentes, C. (2018). Open Science now: A systematic literature review for an integrated definition. *Journal of Business Research, 88,* 428-436.

Weiß, O. (2000). Sport und Gesundheit. *Die Auswirkungen des Sports auf die Gesundheit–eine sozio-ökonomische Analyse*. na.

Wolff, R. F., Moons, K. G., Riley, R. D., Whiting, P. F., Westwood, M., Collins, G. S., ... & PROBAST Group†. (2019). PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. *Annals of Internal Medicine*, *170*(1), 51-58.

Wu, T., He, S., Liu, J., Sun, S., Liu, K., Han, Q. L., & Tang, Y. (2023). A brief overview of ChatGPT: The history, status quo and potential future development. *IEEE/CAA Journal of Automatica Sinica*, *10*(5), 1122-1136.

Zhuang, Y. T., Wu, F., Chen, C., & Pan, Y. H. (2017). Challenges and opportunities: from big data to knowledge in AI 2.0. *Frontiers of Information Technology & Electronic Engineering*, *18*, 3-14.