Machine Learning Applications to Sports Injury: A Review

Hanna Sigurdson¹,³ and Jonathan H. Chan²,³

¹Engineering Science, University of Toronto, Toronto, Canada
²School of Information Technology, King Mongkut’s University of Technology Thonburi, Bangkok, Thailand
³Innovative Cognitive Computing (IC²), King Mongkut’s University of Technology Thonburi, Bangkok, Thailand

Keywords: Machine Learning, Sports Injury, Intrinsic Factors, Extrinsic Factors, Sports Injury Risk Factors, Artificial Intelligence, Literature Review, Sports Psychology.

Abstract: As sports injuries increase in frequency in adolescents, and injuries in professional athletes create a detrimental impact on the sports industry, research surrounding preventing sports injuries becomes more prevalent. The mechanism for sports injury is well defined and includes intrinsic (age, psychology etc.) and extrinsic risk factors (weather, training load etc.), and the inciting event. With the rise of machine learning (ML), a variety of ML techniques have been applied to various sports injury aspects. The purpose of this work is to assess the current applications of ML to sports injury and identify areas of growth by a systematic analysis of applications to each injury element: intrinsic factors, extrinsic factors, and the inciting event. Current underdeveloped areas are identified as: psychological effect, use of extrinsic factors, analysis of the inciting event, and application of the action recognition ability of videos and wearable technology. Future technical applications in these underdeveloped areas should be undergone to expand on and improve sports injury prevention technology.

1 INTRODUCTION

1.1 Sports Injuries

Sports injuries occur across all levels, including children, adolescents, recreational adults of various age groups, and professional athletes. This has a detrimental impact on multiple aspects of society.

Sports injuries in the pediatric and adolescent population are increasing in frequency (Habelt et al., 2011; Kerssemakers et al., 2009). Injuries can cause long term health impacts for young players (Maffulli et al., 2010). Sports injuries can also have a significant effect on the mental health of the individual at all levels of play, affecting self-esteem and overall performance (Arvinen-Barrow & Walker, 2013; Nippert & Smith, 2008).

In professional sports, an injury can have a significant impact on the player’s career, team performance and the sports industry (Brock & Kleiber, 1994; Warnock, n.d.). An informative example of the magnitude of the monetary cost is the $12.4 million U.S. per team that sports injuries cost in the top four professional soccer leagues in 2015 (Guest, n.d.).

Sports injuries across the general population have resulted in an enormous economic cost for direct medical care, rehabilitation, lost wages, and national productivity losses. In 1994 these factors led to an estimated $224 billion U.S. cost in the United States of America (U. Johnson, 2011a). As a result of the abundance of evidence towards the detrimental effects of sports injuries, significant efforts have been placed into understanding the etiology and identifying various risk factors for sports injuries.

1.2 Machine Learning

The recent advancements of machine learning (ML) have resulted in recommendations to implement ML in sports medicine (Beal et al., 2019; Claudino et al., 2019; Martin et al., 2021; Ruddy et al., 2019; Van Eetvelde et al., 2021).

https://orcid.org/0000-0001-6572-2686
https://orcid.org/0000-0002-2384-0462
ML is a branch of artificial intelligence that uses data to build models that make informed predictions (Martin et al., 2021). It has a large variety of applications, one being image classification and analysis, which can be widely applied in the medical field. To train a model, large quantities of high-quality data are required, which translates to requiring an expert understanding of etiology when applied to sports medicine.

The developing technologies in sports injury prediction and other sports areas based on ML can be used to prevent injuries, therefore addressing the negative impacts of sports injuries discussed. This justifies the importance of novel applications and implementation of such technology.

1.3 Purpose

This work aims to assess the current applications of ML to sports injury and identifies areas of growth and opportunities for new applications based on the current understanding of injury risk factors.

2 THE INJURY MECHANISM

The mechanism of injury has been the topic of numerous works that propose models for sports injury (Bittencourt et al., 2016; Meeuwisse, 1994; Meeuwisse et al., 2007; van Mechelen et al., 1992). A common approach to representing the injury mechanism and sports injury aetiology is categorising various injury risk factors into intrinsic and extrinsic, and then including the occurrence of an inciting event. One such example is Meeuwisse et al.’s multifactorial model shown in Figure 1 (Meeuwisse et al., 2007). Of the risk factors for injury that are distant from the outcome there are the intrinsic factors, which are individual physical and psychological characteristics, and the extrinsic factors which are external factors (Lysens et al., 1984). Generally, both intrinsic and extrinsic factors can be classified into “modifiable” and “non-modifiable” factors.

An example list of risk factors is as follows: extrinsic non-modifiable factors are kind of sport, level of sport, position, time of season, and weather; extrinsic modifiable factors are equipment, playing surface, playing time, rules, and time of day; intrinsic non-modifiable factors are age, previous injury, and sex; intrinsic modifiable factors are coordination, fitness level, flexibility, participation in sport-specific training, proprioception, psychological factors, and strength (Habelt et al., 2011).

Finally, once the predisposed athlete has become susceptible there is opportunity for the inciting event, the mechanism of injury that is proximal to the outcome.

Bittencourt et al. agreed that due to the complex nature of sports injury, future research requires movement from isolated risk factor research to injury pattern recognition caused by interactions in a web of determinants (Bittencourt et al., 2016). This provides further argument for application of ML to sports injury prediction, because of its strength in application to multifactorial predictions.

This work leverages the injury mechanism by examining which factors at various stages of injury have been the focus of ML applications.

3 LITERATURE REVIEW

This section analyzes sports injury risk factors by reviewing the current understanding of sports injury etiology at different sources of risk (intrinsic and extrinsic) as well as different stages of injury (both the risk factors and the inciting event). Furthermore, it aims to understand where recently popularized ML techniques have been applied and where they have yet to be explored. A mind map summary of the results from this review is shown in Figure 2.

3.1 Methodology

This work followed a systematic approach to review the literature surrounding ML applications to sports injury, while also addressing risk factors that have been identified for sports injuries. The guidelines followed for this review were:
1. The authors conducted searches on the databases Google Scholar, PubMed, and Scopus.
3. Inclusion and exclusion criteria for sports injury risk factors included that the works must relate to sports, and not solely other injury types. Furthermore, the work cannot focus only on injury rehabilitation, as injury rehabilitation is outside of the scope of this work.
4. Inclusion and exclusion criteria for ML applications included that the work must relate to sports; however, the work can apply to aspects of sports other than injury. The works must address some type of ML application.
5. Abstracts of relevant works were read and if the work remained within the inclusion and exclusion criteria, the work was read and included in the review.

3.2 Intrinsic Factors

Each athlete has personal risk factors that increase their likelihood of sports injury. These are referred to as intrinsic factors and are the subject of the majority of ML applications to sports injury prediction and prevention. Generally, these factors are classified into modifiable and non-modifiable factors, and the following will provide an overview of these factors and the applications of ML previously applied to them.

3.2.1 Intrinsic Factors Summary

Various intrinsic factors have been identified through numerous research studies. Some of these include: nutrition (Close et al., 2019) female gender, age greater than 24 years, a high body mass index, low level of physical fitness at the commencement of a training program, a past history of injury, leg length discrepancy, neuromuscular control, core instability, and many more (Chorba et al., 2010) . This is only a select list and does not include all factors.

Other common factors focused on in this work are: psychological factors, stretching and warm up, training load and fatigue.

3.2.2 Psychological Factors

Examining the psychological impact on sports injury occurrence has been a well-established subject for many sports medicine researchers and suggests a meaningful path to injury prediction and prevention (Heil, 1993).

Williams and Andersen proposed a model of psychological antecedents of sports injury and provided a basis for future models shown in Figure 3 (Andersen & Williams, 1988; U. Johnson, 2011a).
Arvinen-Barrow et al. found significant occurrences of sports injuries impacted by factors such as personality, anxiety, stress response, locus of control, mental and emotional stress, major life events, daily hassles and occurrence of sports injuries (Arvinen-Barrow & Walker, 2013). Some evidence for stressors from difficult relationships or disagreements and inability to cope may predict sports injuries, and discussion of these factors is important before the start of the season (Gould & Petlichkoff, n.d.).

Furthermore, attention to psychology related to injury rehabilitation is very common, likely because injury gives incentive to monitor an athlete’s psychological state (Brewer & Cornelius, 2003; Concannon & Pringle, 2012). Fernandes et al. suggested that there needs to be sufficient athlete support after injury; however, in some situations, psychological intervention may be detrimental and requires educated coaches and rehabilitation supporters (Fernandes et al., 2014). Psychological impact on injury rehabilitation has also been approached from the angle of the effect of “burnout”, which can disturb the regulation of the central nervous system, the autonomic nervous system, and the neuroendocrine system. This effect may manifest in terms of physical symptoms and/or behavioral changes (Ahern & Lohr, 1997).

Among psychological analysis, gender is also considered to be a factor that may influence sports injury. Wiese-Bjornstal et al. observed that perfectionistic, competitive, and compulsive personality factors predicted greater frequency of overuse injury in female collegiate athletes (Wiese-Bjornstal et al., 2015). After extensive literature review, they concluded that gender did play a role in sports injury and is worth further investigation.

3.2.3 Stretching and Warm-up

Stretching and warm-up before and after exercise or sports is widely recommended to prevent injury and improve player performance (Thacker et al., n.d.). Despite this shared standard for athletes at the professional and recreational level, a literature review conducted by Thacker et al. did not find sufficient evidence to make conclusions that stretching and pre-play warm ups could prevent injury, and suggested further investigation in this area (Thacker et al., n.d.). While they did however find some evidence for extremes of inflexibility and hyper flexibility to increase the risk of injury. Particular focus on hamstring muscle injury, however, led Worrell and Perrin to conclude that stretching, hamstring flexibility, warm-up, strength and fatigue all can affect hamstring injury and should be used to help injury prevention (Worrell & Perrin, 1992).

Functional screening tools assessing intrinsic factors of joint laxity, age, decreased range of motion of hip abduction, height (knee ligament injuries), may be good predictors of various injury types. Specifically, ankle injuries could use predictors such as greater strength of the plantar flexors, greater BMI, and postural sway (Dallinga et al., 2012).

3.3 Intrinsic Factors and ML

Of all the applications of ML to sports injuries, use of intrinsic factors to predict future injuries was found to be the most common application. After review, a variety of ML algorithms have been identified as being used for prediction of injury based on intrinsic factors. Generally, the features...
selected can be separated into non-modifiable/longer-term, modifiable/shorter-term, and a combination of both.

### 3.3.1 Non-modifiable and Longer-term Factors

Although works focusing solely on non-modifiable intrinsic factors are rare, they have been able to produce predictive models with Area Under the Curve (AUC) values around 0.65. Studies focusing on unmodifiable risk factors such as sex, knee joint laxity, medial knee displacement, height and other factors such as socioeconomic status were used as predictors in adolescent basketball and floorball players (Jauhiainen et al., 2021). A similar work focused on adolescent athletes (Karuc et al., 2021) in an attempt to respond to growing adolescent sports injury. An application to dental injuries including non-modifiable factors as well as history of dental injury and socioeconomic class also produced meaningful predictive ability (Farhadian et al., 2020).

In adult Australian footballers, non-modifiable factors such as age, stature, mass, primary playing position, and lower limb injury history achieved an 85% accuracy using XGBoost, indicating that these factors can produce relatively accurate injury prediction (Ruddy et al., 2018).

### 3.3.2 Modifiable and Shorter-term Factors

Player fatigue and overuse of the body has been used as an injury predictor, which reflects the literature previously discussed that supports fatigue as an intrinsic factor. Using a Subgroup Discovery approach on 14 elite male volleyball players, risk factors such as fatigue, overuse, sleep, muscle soreness and training exertion produced positive predictive results (Leeuw et al., 2021). A similar application to professional soccer players using predictors of high metabolic rate and sudden decelerations using a Decision Tree predicted 58% of injuries (Rossi et al., 2017), indicating the importance of fatigue measurement in injury prediction.

The motion and mobility of legs or other parts of the body has also been a popular measured value for injury prediction. To predict ACL injuries in basketball players, Taborri et al. used the features of leg stability, leg mobility, and capability to absorb the load after jump measured by inertial sensors and optoelectronic bars and obtained a 96% accuracy using Support Vector Machine (Taborri et al., 2021). Motion assessment (e.g. single leg hop distance, tuck jump assessment etc.) in youth soccer players also obtained reasonable injury predictability using a Decision Tree model (Oliver et al., 2020).

Using wearable technology to predict injury has seen some early implementation; for example, IMUs and surface electromyography (sEMG) electrodes were used to measure predictors such as obliquity of the pelvis, fall of the contralateral pelvis, the extension of the knee, dorsiflexion of the ankle in the initial contact, and less activation of the gluteus medium during the first phase of float in triathletes (Martínez-Gramage et al., 2020). A Random Forest model achieved an AUC of 0.8 for predicting injury from this data, and Random Forest also succeeded when predicting ACL reinjury using fine-grained motion data, with an AUC value of 0.89 (Kim et al., 2019). Future use of wearable technology during movement and during the inciting event could be a future step in predicting and understanding injury.

### 3.3.3 Combination of Modifiable and Non-modifiable Factors

The logical progression for injury prediction is to include both modifiable and non-modifiable intrinsic factors. Physical attributes such as height and weight used in conjunction with physical ability such as strength, flexibility, speed, agility, and endurance in youth soccer players achieved 85% precision using XGBoost (Rommers et al., 2020). The same predictors applied to various NCAA athletes achieved an AUC of 0.79 (Henriquez et al., 2020). From these examples we can see the success of using both modifiable factors and non-modifiable factors that describe the physical condition of the athletes prior to injury.

Another interesting example uses player and team statistics with previous injury statistics in NHL players to produce an AUC performance from XGBoost of 0.956 (Luu et al., 2020). This leaves one to question which features in those statistics indicate injury, as they do not provide a direct correlation to injury in etiology studies. This example points to the value of attempting various data sources even if there is not much evidence initially of the impact of the feature.

Although it may be hypothesized that works combining both modifiable and non-modifiable factors as predictors would produce higher accuracies, the accuracy of each model is highly variable based on previous works. For example, Lopez-Valenciano et al. used a variety of predictors including position, current level of play, dominant leg, age, body mass, stature, body mass index, sleep quality and burnout as well as physical attributes like...
dynamic postural control, lower extremity joint ranges of motion and more (López-Valenciano et al., 2018). However, they only achieved an AUC value of 0.76, which is not an improvement over other models using fewer features. Direct comparison between models including specific features could provide better insight into the preferred risk factors to measure. However, due to the current literature that has highly specified datasets and little broad application across varying datasets, it is difficult to compare models trained on different datasets. Less fine-tuning to specific datasets in future works could be useful to apply models in a broader context.

### 3.3.4 Discussion of Intrinsic Factors

Despite the overwhelming evidence of psychological risk factors, there is a lack of ML applications to these risk factors. This can be hypothesized to be a result of the private nature and difficulty of testing psychological aspects of an athlete.

Many of the works relating to intrinsic factors are limited in the robustness of their datasets, sampling only tens of athletes (Huang & Jiang, 2021; Leeuw et al., 2021; Naglah et al., 2018; Taborri et al., 2021). Some works that used larger datasets with hundreds of athletes could produce higher accuracy values and are more generalized, suggesting that this could be a more useful approach when such data is available (Luu et al., 2020; Rommers et al., 2020). Furthermore, many seem to use arbitrary features, choosing certain intrinsic factors but excluding others even though there is sufficient evidence to support the excluded features in injury prediction. More research on the effect of using a broader scope of predictors is an area that may improve the accuracy of current injury predictive technologies.

### 3.4 Extrinsic Factors

Extrinsic or “external” factors are environmental conditions that increase the risk of sports injury (Andersen & Williams, 1988; Fuller et al., 2016).

An important extrinsic factor is the floor or field conditions. An example of this was concluded in (Olsen et al., 2003), where they found that artificial floors resulted in higher injury risk than wooden floors for female team handball players.

The level of play and type of sport can influence sports injury. An informative example is that lower performance levels in combat sports showed a higher frequency of injuries, and in karate, newer competitors are also significantly more prone to injury (Hammami et al., 2018). Between combat sports there was also a discrepancy in sports injury risk based on the type of combat sport (e.g. highest risk in mixed martial arts).

#### 3.4.1 Training Load

A key component of athlete success is regulation of their training schedule and load. Jones et al. conducted a literature review summarizing the understanding of impact of training load and fatigue on injury and illness (Jones et al., 2017). One of their conclusions verified that during periods of training load intensification, accumulation of training load and acute change in load can increase risk in injury and modifying training load during these times could aid in preventing injury or illness. Although some conclusions were made, they also noted the difficulty to make broad conclusions due to other intrinsic and extrinsic factors such as fitness, body composition, playing level, injury history and age (Jones et al., 2017). Another example case study is prevalence of patellar tendon injuries in elite male soccer players, where multivariate logistic regression showed that high total exposure hours and increased body mass were risk factors (Hägglund et al., 2011).

### 3.5 Extrinsic Factors and ML

Despite the evidence of the effect of extrinsic factors on sports injury risk, there are only a few applications of ML to this aspect of sports injury to the extent of the knowledge of this review.

Some evidence of application is focused on training load, and the impact of playing or training time on injury. One example used GPS tracking data to describe the training load of 26 professional soccer players, and then applied a Decision Tree model to predict 80 % of the injuries with about 50 % precision (Rossi et al., 2018). GPS training data in soccer using an XGBoost model achieved 95 % accuracy (Vallance et al., 2020), indicating the advancements of this method in injury prediction. Similar work using GPS data to assess exertion in rugby players also produced promising results (Thornton et al., 2017), which proves the ability of application across multiple sports. Other works included training hours or time in conjunction with intrinsic factors to predict sports injuries (Naglah et al., 2018; Song et al., 2021).

A sole example of using the playing surface applied a proposed Artificial Neural Network to 21 soccer players, and included proposed epidemiological predictors of interface, shoes, contaminants, damage on the playing surface and the
athlete's movement's nature (Huang & Jiang, 2021). Future inclusion of these features could potentially improve injury prediction.

The inclusion of weather as features during injury prediction is not common but has shown promise: when applied to soccer games (Landset et al., 2017) and ski injuries (Radovanovic et al., 2019). Further use of the weather in predictions for outdoor activities is a promising avenue for future model features.

3.6 Inciting Event

The inciting event is considered the immediate cause of injury and is often the extent of injury knowledge for many people. Due to the short-term nature of the occurrence, very little aspects of the inciting event have seen ML application as compared to the intrinsic factors. This review includes only one work relating to the inciting event.

The early detection of injuries in Major League Baseball Pitchers using video (Piergiovanni & Ryoo, 2019), is an informative example of the power of deep learning to make predictions using computer vision. Piergiovanni et al. used video framing baseball pitchers and applied optical flow (a method used for motion recognition). They then used a Convolutional Neural Network (CNN) trained on the optical flow to binarily classify pitches as an injury or not an injury. Some injuries such as hamstring strains achieved high predictive accuracies at around 0.98, while low accuracy for finger blisters at 0.64. This is an illustrative example of a ML application to further understand and prevent the inciting event.

3.7 Rehabilitation

To narrow the scope of this review, works relating to rehabilitation have been excluded. Some examples of ML applied to sports injury rehabilitation include using visual analysis or other means to assess rehabilitation progress (Ba, 2020; Chen & Yuan, 2021; Edouard et al., 2020; Su, 2019).

3.8 Other Sports Applications

Apart from injuries, recently ML has made significant progress in applications to other aspects of sports (Advancing sports analytics through AI research, n.d.). Some examples of this include tracking players, predicting motion, analyzing shots, understanding strategy and so on, see a mind map summary shown in Figure 4. Although this area has been a topic of discussion in previous reviews, they are limited in their analytical conclusions (Rajš & Fister, 2020) or breadth of research (Richter et al., 2021), and if injury is included they fail to address known injury risk factors (Claudino et al., 2019).

One popular area of research is game outcome prediction, which was excluded from this review due to the large scope of the area and that it is less likely to be translated into future sports injury application than methods such as wearable technology and video analysis. Using player and team statistics, such prediction can achieve around 70 % accuracy, and is currently being developed (Cao, 2012).

The following section provides a summary of the current applications and provides insight on their potential applications to sports injury.

3.8.1 Video Data Analysis

In recent years computer vision advancements have led to success in analysis using video as input data. A main challenge focus has been the identification of actions in video and is an essential step towards injury prediction through video. Recognizing tennis shots (Mora & Knottenbelt, 2017; Skublewskas-Paszkowska et al., 2020), recognizing soccer events (Vats et al., 2020), annotating tennis events (de Castro, 2018), recognizing basketball action (such as pass, shoot, catch, dribble) (Ji, 2020), knee joint moment estimation (W. R. Johnson et al., 2019), and automatic highlight generation (Lee et al., 2020, p.) are applications identified in this review.

The use of video data to track multiple players is an important application of ML to gain more understanding of the influence of players and contact in team sports on injuries. Evidence of this technology is when Tian et al. used SportVU data from the National Basketball Association teams to achieve 69 % accuracy in defensive strategy classification using a KNN model (Tian et al., 2020).

3.8.2 Wearable Technology Data Analysis

The use of inertial measurement units (IMU) to collect data for action recognition is a common application of ML in sports. This provides opportunity for injury prevention, play enhancement and coaching support (McGrath et al., 2020).

Similar to the developments in video, wearables applied to a sports context are mainly focused on activity recognition: general motion (Barshan & Yüksel, 2014), basketball motions (Hu et al., 2020), team handball throws and their speed (van den Tillaar et al., 2021), kayakers’ strokes (Liu et al., 2021), and tennis strokes (Sharma et al., 2017). Various models were applied, for example: Artificial Neural Network, Support Vector Machine (Gourgari et al., 2013),
Random Forest, and Gradient Boosting (van den Tillaar et al., 2021).

Wearable data applied to sports injury has focused on identification of previous injury, for example anterior cruciate ligament injury in rugby players (Tedesco et al., 2020). This leaves an opening for the use of wearable data to predict injuries based on the athlete motion, or to analyze the precursors to the inciting event.

An important observation is the disparity in predictive ability based on skill level, where data collected on professional athletes achieves a higher accuracy which could pose difficulty when applying the same model to novice players. When applied to experienced players, the classification accuracy of basketball action (shooting, passing, dribbling, and lay-up) was 0.85, while only 0.65 for inexperienced players (Hu et al., 2020). This may impact the application of any sort of action or injury recognition to younger or novice players. Despite these players being at risk of injury, the same video or wearable technology data may not be able to achieve high predictive accuracies because of their inconsistency of play and is an area to consider in future works.

Finally, fatigue measured using wearable technology is a useful development that could improve the prediction of sports injury. Predicting fatigue during various exercises such as squats (Jiang et al., 2021), and predicting bicep muscle fatigue (Elshafei & Shihab, 2021) are two examples that predict fatigue of the athlete. As discussed in section 3.2.5, fatigue is a verified risk factor for injury, and this could provide an effective method of measuring fatigue to be used as a short-term predictor of sports injury in future models.

4 CONCLUSIONS

As sports injuries become a growing concern in the adolescent population and adversely affect professional athletes, future application of ML to improve injury prediction and prevention is necessary. This review provides an in-depth understanding of the current areas of deficiency in the ML applications to sports injury. Main areas that are currently the focus of ML application are prediction based on modifiable and non-modifiable intrinsic factors, with some work focusing on extrinsic factors as well, and very little addressing the inciting event. More applications to extrinsic factors, psychological factors, and the inciting event would be meaningful next steps. For example, leveraging advancements in wearable technology and video data analysis in sports could drastically reduce sports injury concern. Future technologies that can more accurately predict and prevent injuries can help achieve the ultimate goal of this work and all works attempting to lessen the impact of sports injuries.

ACKNOWLEDGMENTS

Thank you to Anika Shenoy, Abnash Kaur Bassi, Arjun Sharma, Connor Glossop, Serena Xinran Liu, Shafinul Haque and Andy Wei Liu, all from Engineering Science at the University of Toronto for support and encouragements during the creation of this work.

REFERENCES


