A Survey on the Use of Artificial Intelligence for Injury Prediction in Sports

Salvatore Tedesco*, Sebastian Scheurer, Kenneth N. Brown, Liam Hennessy, Brendan O'Flynn

Abstract— **Artificial Intelligence (AI) could play a significant role in injury prediction in sports due to its capabilities to detect and identify hidden patterns across multi-modal heterogeneous data sources. This paper aims at providing an up-to-date survey of the state-of-the-art in machine learning for injury predictions in sports. Finally, a number of considerations have been also drawn to discuss about the future research challenges required to be tackled to move this field forward.**

I. Introduction

Nowadays, athletic performance analysis in sports, in terms of technique assessment and injury prevention, has experienced a surge in interest amongst the research community [1], and these aspects are seen to be of great benefit to coaches and athletes. In particular, injury prevention has been considered as one of the main priorities in sports due the associated high physical, mental, and economical burden on the athletes, coaches, and clubs. For instance, in one study, more than 50% of subjects in a cohort of elite adolescent athletes reported at least one new injury, while 91.6% reported a 1-year injury prevalence [2]. As indicated in [3], the lower limb accounted for more than one half of all reported injuries. Moreover, a player's recovery and rehabilitation outcome can be affected by individual injury experiences, which can potentially extend the injury process and player return-to-sport [4]. Furthermore, the immediate healthcare costs are substantial. As estimated by FIFA, 30 billion dollars are spent globally every year for injury treatments in soccer [5]. Likewise, it was estimated that, in just a single football season, English Premier League clubs lost between 19 and 26 million U.S. dollars in players' wages due to injuries [5]. As a result, significant investment in technology has been carried out with the goal of preventing injuries based on the principle that an objective and systematic performance monitoring and evaluation of an athlete's activities can be useful to improve athletic performance while minimizing injury risk [6].

It is therefore evident that Artificial Intelligence (AI) and machine learning (ML), which are already disrupting clinical medicine in a number of contexts (e.g. mortality prediction [7], rehabilitation [8], disease diagnosis [9], and many more), could play a hugely important role in this field based on its capabilities to detect and identify hidden patterns from heterogeneous data. Given the limited body of literature carried out on AI for injury predictions in sports, this paper aims at providing an up-to-date survey of the state-of-the-art in the field with the goals of highlighting the few works carried out in literature and providing recommendations on research challenges to be addressed for future studies.

II. Methods

We searched all the peer-reviewed literature available to University College Cork's library via the library's OneSearch discovery portal. We used a two-stage search approach.

In the first stage, we searched for relevant literature review or survey articles by prepending (TitleCombined:(review or survey)) AND to the following base query:

((TitleCombined:(injury)) OR (Abstract:(injury)))
AND ((TitleCombined:(sport OR athlete))
OR (Abstract:(sport OR athlete)))
AND ((TitleCombined:("machine learning"
OR "artificial intelligence" OR predict))
OR (Abstract:("machine learning"
OR "artificial intelligence" OR predict)))

This yielded six hits, two of which [10-11] are relevant. A number of relevant papers mentioned in [10-11] were also considered in this survey.

Following this part, we moved to the second stage of the search. In this stage, we searched the literature with the base query, but limited the search to literature published after May 2018, the end of the period covered in the most recent review [10]. This yielded 122 hits, whose titles and abstracts were screened for pertinence and eligibility. We only considered publications that:

1. predict future injuries of the limbs, back, or head in individual athletes,
2. clearly define the considered injuries,
3. are peer-reviewed,
4. are methodologically sound, and
5. quantitatively report results (e.g., accuracy, F1-score, etc.).

This yielded one additional eligible article [12]. Overall, eight studies have been currently covered in this analysis. Table I lists the results from the pertinent publications. In

This publication has emanated from research supported by a research grant from the Disruptive Technologies Innovation Fund (DTIF) project HOLISTICS funded by Enterprise Ireland (EI). Aspects of this publication have emanated from research conducted with the financial support of Science Foundation Ireland under Grant number 12/RC/2289-P2 INSIGHT, which is co-funded under the European Regional Development Fund.

S. Tedesco, and B. O'Flynn are with the Tyndall National Institute, University College Cork, Cork, Ireland.

S. Scheurer, and K.N. Brown are with the Insight Centre for Data Analytics, Dept. Computer Science, University College Cork, Cork, Ireland.

L. Hennessy is with Setanta College Thurles Enterprise Centre, Jimmy Doyle Rd, Thurles, Co. Tipperary, Ireland.

* Corresponding author. e-mail: salvatore.tedesco@tyndall.ie

Section III we briefly discuss each of the pertinent papers from the table.

III. Results and Discussion

In 2019, Claudino et al. [10] conducted a systematic review of the literature on the use of Artificial Intelligence (AI) for injury risk assessment and performance prediction in competitive team sports covering scientific journals up to May 2018. Of the fifty-eight papers that met their inclusion criteria — with a pooled sample of athletes (97% male, 25 ± 8 years; 3% female, 21 ± 10 years) —fifteen studies (≈ 26%) relate to injury risk. The 15 studies assessed injury risk in soccer, basketball, handball, as well as American and Australian football. However, only four of these relate to muscle injury prediction, with the remainder concerned with diagnosing injuries, identifying injury risk factors, or estimating their relative importance.

Also in 2019, Beal, Norman, and Ramchurn [11] published a high-level survey of the literature on how AI is used in team sports for match outcome prediction, strategic and tactical decision-making, fantasy sport games, and managing injuries. However, when it comes to applying AI to the wide range of data available to most modern professional sport teams they note that:

"These [GPS-tracker] data alongside the historical medical data that are collected by physios and club doctors can give a feature set for players who [sic.] has yet to be studied by the AI community."

They do, however, cite sports medicine research that shows that higher (relative) workloads are associated with a greater injury risk in soccer, American football, and rugby. Specifically, Hulin et al. [13] have shown that higher acute:chronic workload ratios (ACWRs[1]) are associated with higher injury risks in Rugby League players. They conclude that higher workloads can have either positive or negative influences on injury risk. In particular, players with a high chronic workload are considered more resistant to injury compared with players with low chronic workload; indeed, athletes with moderate-low through moderate-high (0.85–1.35) ACWRs were less resistant to injury when subjected to "spikes" in acute workload (e.g., very-high ACWRs ≈1.5).

Looking at each individual paper, Thornton et al. [14] used various training load measures (i.e., session RPE × session time, GPS distance in different running intensity bands) as inputs to a random forest to predict rugby player's risk of non-collision injuries. The individualised models (e.g., one model per athlete) achieved a mean area under the receiver-operator characteristic curve (AUC) of 0.65 ± 0.05 across all 25 athletes.

Ruddy et al. [15] used age, hamstring injury history, and eccentric hamstring strength as features to predict hamstring injury in elite Australian footballers. Within-year models were built for two seasons (2013 and 2015), as well as between-year hamstring injury prediction models. Median AUC was 0.58 for the 2013 models and 0.57 for the 2015 models, while a median AUC of 0.52 was obtained for the between-year models. Although some iterations of the models achieved near perfect prediction, the variance in the estimated AUC highlighted the fragility of the data collected. The authors concluded that the considered risk factors could not be used to identify athletes at an increased risk of hamstring injury consistently [15].

López-Valenciano et al. [16] investigated models for predicting lower extremity muscle injuries and obtained an AUC of 0.747 by using a SmooteBoost technique with cost-sensitive ADTree as base classifier, with variables encompassing personal (demography, history of injury), psychological (sleep quality, burnout), and neuromuscular risk factors (e.g., dynamic postural control isometric hip abduction and adduction strength, joints range of motion, core stability, isokinetic knee flexion and extension strength). The prediction model showed moderate accuracy for identifying professional soccer and handball players at risk of general muscle injuries. In [17], the same authors developed a specific model for hamstring injury prediction using the same risk factors. The prediction model showed moderate to high accuracy. The best model was obtained via SmooteBoostM1 with cost-sensitive ADTree as base classifier which reported AUC of 0.837.

While [16-17] only considered risk factors obtained from a single screening session, Rossi et al. [18] obtained comparable results (AUC = 0.76) by relying on training workload (kinematic, metabolic and mechanical features) from each training session during the season via GPS devices, thus guaranteeing a constant and individualized monitoring of each training session workload during the season.

Rommers et al. [12] predicted injuries in elite youth soccer players with an XGBoost model from pre-season anthropometric measurements (height, weight, and sitting height), and physical fitness and motor coordination test batteries. The Shapley Additive Explanations (SHAP) analysis revealed that the five most important features were higher estimated age at peak height velocity (PHV), higher body height and leg length, lower fat percentage and average performance on the standing broad jump (SBJ). On a 20% holdout sample of 143 players, the approach predicted injuries with F1-score of 85%.

Carey et al. [19] investigated the possibility to use training load data (quantified using GPS and inertial sensors, as well as player perceived exertion ratings) to predict injuries in elite Australian footballers. Absolute and relative training load metrics were calculated daily for each player. Injury prediction models were built for non-contact, non-contact time-loss, and hamstring specific injuries. For models of non-contact and non-contact time-loss injuries, performance tended to be limited (AUC<0.65), while hamstring-specific injury models achieved better results (AUC=0.76) obtained with a multivariate logistic regression. Moreover, injury prediction models built using data from a single club showed poor predictive scores when tested on previously unseen data, highlighting possible overfitting due to the fragility of the data; however, predictive performance improved with increasing quantity of data. Even though the authors considered training load as possibly an important risk factor in injury prediction, due to overfitting and fragility of the data, they did not suggest

[1] ACWR is defined as $ACWR_t = L_{t-7}/L_{t-28}$, where L is the player's load and t the current day

the developed model as a possible daily decision tool for real-world practice. Instead, they recommended future studies focus on specific injury types, the adoption of additional variables beside training load, and highlighted the need for collaboration and larger cohorts rather than single team data collections.

Finally, training loads were also used by Jovanovic [20], however with overall poor results in terms of injury prediction.

TABLE I. SUMMARY OF PAPERS USING AI TO ESTIMATE/PREDICT ATHLETES' INJURY RISK

Reference	Sport	Data	Injury Location	Injury Rate	N	Best Model(s)	Metric	Performance
Ruddy et al., 2018 [15]	Australian football	Pre-season screening (eccentric hamstring strength, demographic and injury history)	Hamstring	53 cases	362	Naïve Bayes	AUC	0.54 (between-year)
López-Valenciano et al., 2018 [16]	Soccer & handball	Pre-season screening (personal questionnaire, psychological questionnaires (sleep/burnout), and neuromuscular assessment: dynamic postural control, isometric hip abduction and adduction strength, range of motion, core stability, isokinetic knee flexion and extension strength)	Lower extremity muscles	32 cases	132	SmooteBoost with cost-sensitive ADTree	AUC	0.747
Rossi et al., 2018 [18]	Soccer	Demographic and injury history as well as training load-related features extracted from a 10 Hz GPS and a 100 Hz inertial sensor worn over one season	Non-contact injuries	23 cases	26	Decision Tree	AUC	0.76
Thornton et al., 2017 [14]	Rugby	Internal and external training load (RPE scale, sRPE scale, and features extracted from a 5 Hz GPS) obtained over three seasons	Non-contact injuries	156 cases	25	Random Forest	AUC	0.64
Rommers et al., 2020 [12]	Soccer	Pre-season anthropometric measurements (height, weight, and sitting height) as well as test batteries to assess motor coordination and physical fitness (strength, flexibility, speed, agility, and endurance)	Overuse and acute injuries	368 cases	734	Extreme Gradient Boosting (XGBoost)	F1	0.85 (while classifying overuse vs acute injuries had a 0.78 performance)
Ayala et al., 2019 [17]	Soccer	Same as [16]	Hamstring	18 cases	96	SmooteBoostM1 with cost-sensitive ADTree	AUC	0.837
Carey et al., 2018 [19]	Australian football	RPE as well as training load-related features extracted from a 10 Hz GPS and a 100 Hz inertial sensor worn over three seasons	Non-contact injuries and hamstring specific	388 (non-contact injuries) / 49 (hamstring specific)	75	Random Forest (non-contact injuries) / Logistic Regression (hamstring specific)	AUC	< 0.65 (non-contact injuries) / 0.76 (hamstring specific)
Jovanovic, 2018 [20]	N/A	Day-to-day training load proprietary metrics collected over two seasons	Hamstring	25 cases	52	Logistic Regression	AUC	< 0.65

N: number of athletes, AUC: Area Under the Curve

IV. FUTURE CHALLENGES

This survey provides an up-to-date overview on the state-of-the-art in ML for injury predictions in sports. Although, a few literature reviews were published in the field [10-11], they only covered studies until 2018 and did not focus specifically on AI for injury prediction, thus an up-to-date overview was lacking.

Despite the large academic interest in the field, only a limited amount of work has been published that leverages the potentially positive effects of machine learning. This may be due to the logistic problems associated with carrying out large prospective studies, which require collecting data from ideally hundreds of athletes over several sport seasons. To this purpose, the development of collaborative research clusters and consortia for the support of large-scale pilots, which could collect large datasets and make them openly available to the research community, is essential to accelerate the development of high-performing data-hungry AI models for this problem. As a consequence, the creation of common standards in terms of protocols (e.g., test batteries to be adopted, variables collected), device interoperability (to allow the exchange and sharing of a multitude of heterogeneous data sources), security (to preserve athletes' privacy and confidentiality), and ethical principles are a must to allow the implementation of this data collection on large cohorts. Likewise, to guarantee that the AI models developed are unbiased and do not show performance drops when applied on unseen data, best practices need to be identified and disseminated across industry and academia [21].

The works discussed in this survey generally considered variables that either involved lab tests at pre-season (e.g., flexibility, strength) or adopted features extrapolated from wearable data. Only [19] considered a mix of both approaches. As confirmed in [11], taking advantage of wearable devices in this field is still limited. Across the wearable space, GPS and inertial sensors worn on the upper body for training load estimation still seem the most considered option, despite the huge range of biomedical sensors currently investigated by researchers for health biomarkers in sport (i.e., biochemical markers, heart rate, muscle oxygen saturation, etc.) [22]. Despite a number of companies that provide wearable products for continuous athlete monitoring and injury prevention [22-23], those products generally lack AI capabilities and, moreover, there is a lack of validation studies of most of those products in relevant sport cohorts. An increasing number of companies (i.e. Zone7 [24]) are currently raising funding to deliver products combining wearables, video technology, and AI for injury prediction for elite athletes in a variety of team sports. Those systems are, however, cloud-based which may have implications on the data security, system infrastructure, communication latency, network bandwidth, and system scalability. Thus, the use of edge analytics on those wearable systems has recently gained much interest in literature because of the possibility to offer data processing services as close as possible to the data source and to compute part of these data locally, instead of relying on cloud services. The edge computing paradigm offers therefore several benefits on the system infrastructure, helps to main the privacy of the athletes, and allows a fast and real-time decision support system which is adaptive and relies on personal data [25]. Finally, requirements related to the explainability of the model developed (e.g., models which are not black-boxes and whose predictions are interpretable and easy-to-understand by humans while maintaining a high level of learning performance), as well as the introduction of prescriptive analytics (which not only predict the probability that an event may occur but also recommend the best decisions and highlight their implications), are still in their infancy in this field [26]. A summary of the described challenges is shown in Figure 1.

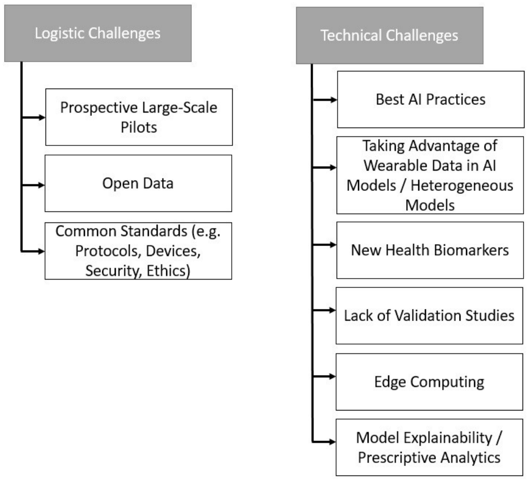

Figure 1. Summary of logistic and technical challenges

In conclusion, while machine learning for injury prediction has become an established concept that is gaining attention in the research community, very few studies have been carried out yet, as shown in this survey. We discussed the challenges that hamper the use of AI for injury prediction in sports, and highlighted some promising areas for future research to move this field forward.

REFERENCES

[1] V. Camomilla, E. Bergamini, S. Fantozzi, G. Vannozzi, "Trends supporting the in-field use of wearable inertial sensors for sport performance evaluation: A systematic review," *Sensors*, 18, 873, 2018.

[2] P. von Rosen, A. Heijne, A. Frohm, C. Friden, A. Kottorp, "High injury burden in elite adolescent athletes: A 52-week prospective study," *J Athl Train*, vol. 53, no. 3, pp. 262-270, 2018.

[3] J.M. Hootman, R. Dick, J. Agel, "Epidemiology of collegiate injuries for 15 sports: Summary and recommendations for injury prevention initiatives," *J Athl Train*, vol. 42, no. 2, pp. 311-319, 2007.

[4] G.P. Murphy, R.B. Sheehan, "A qualitative investigation into the individual injury burden of amateur rugby players," *Physical Therapy in Sport*, vol. 50, pp. 74-81, 2021.

[5] M. Mohib, N. Moser, R. Kim, M. Thillai, R. Gringmuth, "A four year prospective study of injuries in elite Ontario youth provincial and national soccer players during training and matchplay," *J Can Chiropr Assoc*, vol. 58, no. 4, pp. 369-376, 2014.

[6] E.C. Lee, M.S. Fragala, S.A. Kavouras, R.M. Queen, J.L. Pryor, D.J. Casa, "Biomarkers in sports and exercise: Tracking health, performance, and recovery in athletes," *J Strength Cond Res*, vol. 31, no. 10, pp. 2920-2937, 2017.

[7] S. Tedesco, et al., "Investigation of the analysis of wearable data for cancer-specific mortality prediction in older adults," 2021 43rd Annual Int Conf IEEE Eng Med & Biol Soc (EMBC), pp. 1848-1851, 2021.

[8] S. Tedesco, et al., "Motion sensors-based machine learning approach for the identification of anterior cruciate ligament gait patterns in on-the-field activities in rugby players," *Sensors*, vol. 20, 11, 3029, 2020.

[9] F. Jiang, et al., “Artificial intelligence in healthcare: past, present and future,” *Stroke Vasc Neurol*, vol. 2, no. 4, pp. 230-243, 2017.

[10] J.G. Claudino, D. de Oliveira Capanema, T. Vieira de Souza, J. Cerca Serrão, A.C. Machado Pereira, G.P. Nassis, “Current approaches to the use of artificial intelligence for injury risk assessment and performance prediction in team sports: A systematic review,” *Sports Medicine – Open*, vol. 5, no. 28, 2019.

[11] R. Beal, T.J. Norman, S.D. Ramchurn, “Artificial intelligence for team sports: A survey,” *The Knowledge Engineering Review,* vol. 34, Cambridge University Press.

[12] N. Rommers, et al., “A machine learning approach to assess injury risk in elite youth football players,” *Medicine and Science in Sports and Exercise*, vol. 52, no. 8, pp. 1745-1751, 2020.

[13] B.T. Hulin, T.J. Gabbett, D.W. Lawson, P. Caputi, J.A. Sampson, “The acute:chronic workload ratio predicts injury: High chronic workload may decrease injury risk in elite rugby league players,” *British Journal of Sports Medicine*, vol. 50, no. 4, pp. 231-236, 2016.

[14] H.R. Thornton, J.A. Delaney, G.M. Duthie, B.J. Dascombe, “Importance of various training-load measures in injury incidence of professional rugby league athletes,” *Int Journal of Sports Physiology & Performance*, vol. 12, no. 6, pp. 819-824, 2017.

[15] J. Ruddy, et al., “Predictive modeling of hamstring strain injuries in elite Australian footballers,” *Medicine and Science in Sports and Exercise*, vol. 50, no. 5, pp. 906-914, 2018

[16] A. López-Valenciano, et al., “A preventive model for muscle injuries: A novel approach based on learning algorithms,” *Medicine and Science in Sports and Exercise*, vol. 50, no. 5, pp. 915-927, 2018.

[17] F. Ayala, et al., “A preventive model for hamstring injuries in professional soccer: Learning algorithms,” *Int Journal of Sports Medicine*, vol. 40, no. 5, pp. 344–353, 2019.

[18] A. Rossi, L. Pappalardo, P. Cintia, F.M. Iaia, J. Fernàndez, D. Medina, “Effective injury forecasting in soccer with GPS training data and machine learning,” *PLoS One*, vol. 13, no. 7, 2018.

[19] D.L. Carey, K. Ong, R. Whiteley, K.M. Crossley, J. Crow, M.E. Morris, “Predictive modelling of training loads and injury in Australian football,” *Int Journal of Computer Science in Sport*, 17, 1, 49-66, 2018.

[20] M. Jovanovic, “Predicting non-contact hamstring injuries by using training load data and machine learning models,” Technical report, University of Belgrade, Serbia, 2018.

[21] https://www.fda.gov/medical-devices/software-medical-device-samd/good-machine-learning-practice-medical-device-development-guiding-principles

[22] D.R. Seshandri, et al., “Wearable sensors for monitoring the physiological and biochemical profile of the athlete,” *npj digital medicine*, vol. 2, no. 72, 2019.

[23] S. Tedesco, et al., “A wearable system for the estimation of performance-related metrics during running and jumping tasks,” *Applied Sciences*, vol. 11, no. 11, pp. 5258, 2021.

[24] https://zone7.ai/

[25] K. Bierzynski, et al., “AI at the Edge,” 2021 EPoSS White Paper, 2021.

[26] S. Tedesco, “Smart wearable systems for health and wellness in sports, aging, and rehabilitation,” University College Cork, 2022.