

Injury Prediction in Sports: A survey on machine learning methods

Lejia Wang

Received February 05, 2024

Accepted May 03, 2024

Electronic access May 31, 2024

Aims: Emerging machine learning algorithms and models are creating possibilities for consistent sports injury prediction. Deep learning is just one type that is becoming more applied to this aspect of sports medicine as the years progress. Due to the number of new ways of applying these algorithms, this study seeks to analyze the state-of-the-art of machine learning in sports injury prediction through conducting a systematic review.

Methods: A search on PubMed was conducted in June 2023 using key words such as “predictive modeling”, “sports” and “machine learning” and articles were selected using inclusion/exclusion criteria. Certain criteria such as the studies being original ones or being published after 2015 were also considered.

Results: The majority of the studies had a sample size of less than 1000 and around 100. Of these studies analyzed, lower limb injuries were the most common. Interestingly, there were more deep learning studies in the search between 2020 and 2023 than between 2015 and 2020 with most of them presenting higher Area Under the Curve (AUC) values than others. The number of Random Forest and/or Support Vector Machine based studies were the same in both time periods. Cross-validation was the most used training method.

Conclusion: Machine learning-based sports injury prediction models have provided new possibilities for clinical application. The state-of-the-art consists of more ideal values of metrics used to judge the accuracy as well as reliability on these new machine learning models which creates excitement for the future of this field.

Keywords: Machine Learning, Sports, Injury, Prediction, Deep Learning, Random Forest, Support Vector Machine

Introduction

Sports-related injuries occur due to various risk factors with the most prevalent ones being anthropometric attributes or athlete qualities determined through specialized tests, which can then be analyzed to prevent injuries. Thus, through understanding the relationship between these risk factors in relation to the ultimate injury, these injuries can potentially be predicted and prevented. Advanced abilities to make more complex predictions have also been explored more thoroughly through newer algorithms such as deep learning. This is especially useful as it means that it can be applied in injury prediction within sports.

Contact sports and others requiring high levels of intensity are common grounds for injuries. The damage brought upon by them can impact various aspects of an athlete, including their sports career, the coordination of the team or even their life outside of athletics¹. A study conducted on university-level athletes presented an average of 2.28 injuries per athlete with a prevalence of 91% and an overall “lifetime” occurrence amounted to 67.1%². Among professional athletes, muscle injuries are one of the most major challenges and can account for up to one-third of all sports-related injuries³. Short term effects of muscle injuries are play downtime and swelling, but long-term effects can create a great burden on the athlete in the

form of chronic illnesses, such as osteoarthritis⁴. Essentially, no matter if the injury is muscle-related or not, the negative implications are leading issues in athletes and even impact their mental health and academic performance⁵.

Within the past few years, machine learning (ML), a subset of artificial intelligence (AI), has been growing towards vast fields of application, including sports⁶. Sports betting is another way Machine Learning is used in sports. However, the objective of this literature review is solely to investigate and review the role of state-of-the-art ML applications in sports injuries and assess the potential contributions and limitations of these methods in the improvement of injury prediction. The expected contributions are to address, shed light and provide an assessment on this question in using ML.

While various literature reviews have already been published regarding the existing sports injury prediction research using ML, there are only a few and they are not up-to-date. One prior literature review⁷ provided an analysis of ML methods with their effectiveness in sports injury prediction and prevention, specifically with 11 studies meeting their inclusion/exclusion criteria. This study uses this prior literature review as a reference to show state-of-the-art progressions from the lack of ML application along with limited uses of deep learning and unexplored ML algorithms. Notably, it was published in March 2020,

so this study will try to replicate the attributes and methods of conduction of that article to maintain consistency and attempt to find unbiased differences.

Machine Learning Background

Some of the ML algorithms that have translated into state-of-the-art predictions are logistic regression, random forest, support vector machine and deep learning. Logistic regression outputs probability when given input variables, so it is easy to train but lacks the ability to recognize complex relationships⁸. Random forest (RF) is made up of multiple decision trees and summarizes the results of them into one. It is useful because it can be used for both classification and regression⁹. Support vector machine (SVM) separates data and classifies, however, it is lacking when the data has more noise¹⁰. Deep learning is a neural network made of numerous layers. It is most notably used for its ability to learn patterns directly from the data¹¹. The process of using machine learning algorithms typically begins with splitting data into a testing set and a training set to train the model, and then the algorithm is implemented. However, it depends on the algorithm. In deep learning, there is training, validation and testing and in logistic regression, there is only training and testing. Then, features are extracted before data splitting to determine the most impact risk factors for the desired results, which in this case is the highest predicted and accurate value. For deep learning, however, the most impactful features cannot be known, which is why it is called a black box. There were different metrics being reported to convey the results of the performance. The most used one was Area Under the Curve (AUC), but some other metrics used include precision, recall, specificity, sensitivity, Root Mean Squared Error (RMSE) and Brier scores. An AUC value below 0.7 represents poor predictive performance, a value between 0.7 and 0.8 indicates a fair score, 0.8-0.9 is a good score for the model and an AUC value of more than 0.9 is excellent. Then, as a part of the training strategy, the outcomes are tested to see if the results are consistent and can be relied upon. Pre-processing techniques can reduce the number of predictor candidates before training and thus, reduce overfitting risk. They can also be applied for other purposes such as normalizing predictor variables before training the ML algorithm or applying other more complex transformations to remove variability. Normalization is a preprocessing technique that prevents features from dominating one another. One recurrent and comprehensive training strategy that work together with preprocessing is called cross-validation, which involves the split training and testing set mentioned beforehand. Cross-validation is a resampling method where a result is provided after the model is trained on the training set and tested on the testing set. It is useful in understanding the model's generalization ability. Postprocessing techniques include bagging or boosting which combine multiple methods to improve overall

performance.

Through preprocessing and selecting the predictors, the models are more likely to miss relevant things but avoid overfitting with few predictors. The models are also more effective to use this method if there is a known or backed up reason for which predictors are the most impactful in the result. On the other hand, by training the algorithm first, it is possible to capture hidden trends, however, there is a higher chance of overfitting. In machine learning, overfitting is when the training of the algorithm is done too much, and as a result, noise is captured. Furthermore, larger sample studies connect with deep learning or technology. This is because it helps reduce overfitting and overgeneralization and is needed for the many parameters in its nature.

Methods

This study uses the same inclusion/exclusion criteria as⁷ in order to maintain consistency of the types of results and allow for comparisons of time periods. The search was conducted on PubMed under the following criterion: (“deep learning” OR “artificial intelligence” OR “machine learning” OR “neural network” OR “neural networks” OR “support vector machines” OR “nearest neighbor” OR “nearest neighbors” OR “random forest” OR “random forests” OR “trees” OR “elastic net” OR “ridge” OR “lasso” OR “boosting” OR “predictive modeling” OR “learning algorithms” OR “bayesian logistic regression”) AND (“sport” OR “sports” OR “athlete” OR “athletes”) AND (“injury” OR “injuries”). The inclusion criteria were as follows:

1. Original studies investigating the use of ML in predicting sports injuries published in peer-review journal;
2. English-language studies.

The exclusion criteria were as follows:

1. articles published before 2015
2. not being sport-specific or covering injury prediction
3. conference or meeting abstract.

The study selection was conducted following a 2-step process: the studies were first shortlisted based on their titles and abstracts and according to the on the inclusion and exclusion criteria, and the full text was subsequently analyzed for final inclusion.

Results

The keywords search conducted in mid-June of 2023 resulted in 523 entries, where 25 out of them met the selection criteria.

[Summarized chart on all studies analyzed in this literature review. Covers the title, injury, sport, ML technique or model,

Table 1 Summarized ML Algorithms Prediction of Injuries Study Characteristics

Article Title	Injury	Sport	ML Technique/Model	Predictors (Number, Type, ...)	Performance (Metrics, Performance VS. Baseline)
Predictors of Ulnar Collateral Ligament Reconstruction in Major League Baseball Pitchers ¹²	Ulnar collateral ligament injuries	Major League Baseball (MLB)	Binary logistic regression, naive Bayes, support vector machine, binary linear regression	14 predictor variables	The binary linear regression model: statistically significant (P = 0.01), correctly classified 66.8% of cases Naive Bayes classification accuracy 72% Support vector machine classification accuracy: 75%
Importance of Various Training-Load Measures in Injury Incidence of Professional Rugby League Athletes ¹³	Time-loss, soft-tissue, overuse injuries	Professional rugby (National Rugby League competition)	Generalized estimating equations (GEE) model and random forest	21 predictor variables	GEE Models: Adjustables - QIC of 566.5 and = 0.001 to 0.091. Hit-up forwards - QIC of 441.7 and p = 0.006 to 0.138. Outside backs QIC of 406.6 and = 0.092 to 0.225. Wide-running forwards QIC of 410.5 and p = 0.068 to 0.830 Random forest Mean (SS) ROCE for all the groups around 0.6 and 0.7.
Predictive Modeling of Hamstring Strain Injuries in Elite Australian Footballers ¹⁴	Hamstring strain injuries	Australian Football	Supervised learning techniques (Naive Bayes, Logistic regression, Random forest, Support vector machine, Neural network)	9 predictor variables	-0.57 to 0.59 media AUC for any of the models
Effective injury forecasting in soccer with GPS training data and machine learning ¹⁵	Non-contact injuries	Soccer	Decision tree algorithm	55 features	Decision tree classifier DT has recall = 0.80+0.07 and precision = 0.50+0.1 on the injury class, can predict 80% of injuries and can label a training session as an injury in 50% of the cases
A Preventive Model for Muscle Injuries: A Novel Approach based on Learning Algorithms ¹⁶	Lower extremity muscle injuries (MUSIN))	Professional soccer and handball	Decision tree algorithms (with Random tree as well), ensemble learning algorithms	52 features	ADTree achieved best performance in most of analyzed methods (AUC values of 0.6-0.7)
A Preventive Model for Hamstring Injuries in Professional Soccer. Learning Algorithms ¹⁷	Hamstring strain injury	Professional soccer	Decision tree algorithms: J48, ADTree and SimpleCart.	Modifiable and unmodifiable risk factors Personal risk factors, psychological risk factors, neuromuscular risk factors	AUC of 0.837 for best performing model
On-field player workload exposure and knee injury risk monitoring via deep learning ¹⁸	Non-contact knee trauma	N/A	CaffeNet CNN (convolutional neural network)	59 features	The strongest mean KJM correlation is for the left stance limb during sidestepping (r= 0.9179) which is also during sidestepping (r = 0.8168)
A Machine Learning Approach to Assess Injury Riskin Elite Youth Football Players ¹⁹	All kind	Football (soccer)	Gradient boosting (XG-Boost)	29 predictors from preseason test results	Metrics: Precision, recall (sensitivity), accuracy (f1 score) Extreme gradient boosting model: Precision of 84%, recall of 83%, and an f1 score of 83% in the training dataset. On the test data, the precision, recall and f1 scores were all 85% (reasonable accuracy and sensitivity) Classifying injuries - reasonably accurate in classifying injuries correctly

Using machine learning to improve our understanding of injury risk and prediction in elite male youth football players ²⁰	Non-contact lower-limb injuries	Elite football	Decision trees: (J48ccn), an alternating decision tree (ADT) and a reduced error pruning tree (REPTree)	6 risk factors	Logistic regression on categorical data: For the univariate analysis, all variables reported an AUC 0.57. With no collinearity, a multivariate analysis offered prediction with AUC of 0.687. Best performing decision tree model was with the bagging ensemble method on the J48con decision tree base classifier. Cross-validation gave an AUC of 0.663 (model correctly classified 74.2% of non-injured players and 55.6% of injured players)
Machine Learning Outperforms Logistic Regression Analysis to Predict Next-Season NHL Player Injury: An Analysis of 2322 Players From 2007 to 2017 ²¹	All kind	Hockey	Random forest, K nearest neighbors, Naive Bayes, XGBoost, Top Three Ensemble	35 features for nongoalie cohort and 14 features for goalie cohort	XGBoost had the highest AUC of 0.948 XGBoost model predicted next-season injury with an accuracy of 94.6% (SD, 0.5%), and an accuracy of 96.7% (SD, 1.3%) for goalies
Machine Learning to Predict Lower Extremity Musculoskeletal Injury Risk in Student Athletes ²²	Musculoskeletal injuries	Division 1 NCAA sports: Basketball, Men's football, Soccer and Women's Volleyball	Random forest	50 physical metrics and demographic data	Initial validation strategy of test/train split ROCAUC accuracy was 79.02%, secondary validation with k-fold cross validation of an average ROC AUC of 68.90%
Basketball Sports Injury Prediction Model Based on the Grey Theory Neural Network ²³	Ankle and knee injuries	Women's basketball	Grey neural network	Improved unequal interval model	Predicted results of most basketball injuries are close to the actual results, SO the network's injury prediction based on gray is reliable.
Injury Prediction in Competitive Runners With Machine Learning [23]	Unspecific running-related injuries	Competitive running	Extreme Gradient Boosting, or XGB boost	Objective data from a global positioning system and subjective data about the exertion and success of the training. 22 aggregate features	Average AUC scores of 0.729 and 0.724 for validation and test sets of day approach, and AUCs of 0.783 and 0.678 for the validation and test sets of the week approach.
Combining Inertial Sensors and Machine Learning to Predict vGRF and Knee Biomechanics during a Double Limb Jump Landing Task ²⁴	Anterior crucial ligament	N/A	Support vector machines (SVMs), artificial neural networks (ANNs), and generalized linear models	14 predictor variables	Across vGRF, KFA, KEM, and KPA, both multiple feature models had an RMSE smaller than the clinical difference, creating confidence in their use. Both single feature models had an RMSE larger than the clinical difference, limiting their use. Although the R2 for the KEM and KPA models did not exceed the 0.8 for high accuracy, the error also did not exceed the expected differences between ACLR and healthy control limbs, thus the models still have clinical value.
New Machine Learning Approach for Detection of Injury Risk Factors in Young Team Sport Athletes ²⁵	Lower extremity injuries	Basketball and floorball	Random forest and 1-l regularized logistic regression	For random forest, 12 consistent injury predictors For L1-regularized logistic regression, 20 consistent injury predictors	For random forest, the mean AUC-ROC value was 0.63 and 0.94 for the training data (values of real responses higher than the randomized, mean AUC-ROC 0.48) confirms significance For logistic regression, the mean AUC-ROC value was 0.65 and 0.76 for the training data (values of with real responses higher than the randomized ones, mean AUC-ROC 0.50)

Predictive Modeling of Injury Risk Based on Body Composition and Selected Physical Fitness Tests for Elite Football Players ²⁶	All kind	Football (soccer)	Linear regression models (Classic regression models (OLS), shrinkage regression, stepwise regression, LASSO)	22 independent variables	For elastic net regression, its prediction error was small (RMSE = 0.633) and the amount of predictors was decreased due to the method's characteristics. The use of shrinkage models (Ridge, LASSO, and elastic net) caused quick decrease in error or (an improvement in the model's predictive ability). The Ridge model was the best-performing model for predicting injuries (RMSE error was 0.698)
Predicting ACL Injury Using Machine Learning on Data From an Extensive Screening Test Battery of 880 Female Elite Athletes [27]	Anterior cruciate ligament	Handball and softball	Random forest, L2-regularized logistic regression, and support vector machines (SVMs) with linear and nonlinear kernel	Player baseline characteristics, elite playing experience, history of any previous injuries and measurements of anthropometrics, strength, flexibility, and balance.	Linear SVM (without any imbalance handling) obtained an AUC-ROC value of 0.63 (highest mean). With all 4 classifiers, great differences between minimum and maximum AUC-ROC values during repetitions because of random cross-validation splits. Training AUC-ROC values were very high for random forest and SVMs, but with logistic regression, control of overfitting was better.
Detecting Injury Risk Using Machine Learning Models in Elite Women's Pathway Cricket ²⁷	N/A	Cricket	Decision tree and random forest	Traditional algorithm decision tree: 1064 observations from 47 input variables. Best-performing random forest model: 1,064 observations from 47 input variables.	Decision tree: poor overall probability of accurately predicting injury with the training data (56% for each rule). On the testing data set (30% of the data randomly split), the conditional algorithm performed poorly (AUC of 0.66, but still slightly better than the traditional algorithm (AUC of 0.57)). Best performing random forest model: On the testing data set, conditional algorithm (AUC of 0.72) performed poorly, but still slightly better than the traditional algorithm (0.65)
Impact of Gender and Feature Set on Machine-Learning-Based Prediction of Lower-Limb Overuse Injuries Using a Single Trunk-Mounted Accelerometer ²⁸	Lower-limb overuse injury (LLOI)	Dance, track and field, gymnastics, swimming, basketball, handball, soccer, and volleyball	Logistics regression, support vector, trees	Basic characteristics and tri-axial acceleration measurements over the L3 to L5 spinal segments while performing the running Cooper test.	Logistic regression (LR) model was best-performing in terms of AUC score (mean AUC score of 0.645, 0.056). (used entire set of features), mean Brier score of 0.19040.021. Second highest is for the models using statistical features, lowest-performing models in terms of AUC only using sports-specific features. Support vector machine models: lowest mean Brier score (best results) for the female-specific, male-specific, and all-data models no matter the feature type or amount
A machine learning approach to identify risk factors for running-related injuries study protocol for a prospective longitudinal cohort trial ²⁵	Unspecific running-related injuries	Competitive running	Deep Gaussian Covariance Network (DGCN)	Internal and external characteristics	Deep Gaussian Covariance Network (DGCN)
A deep learning-based approach to diagnose mild traumatic brain injury using audio classification [33]	mBTI (mild traumatic brain injury)	Rugby	Bidirectional long short-term memory attention (Bi-LSTM-A) deep learning model	38 different vocal features	Little-to-no overfitting in the training process, so the training process on observed data would allow the model to generalize well for classifying unseen data (good reliability). Overall accuracy of 89.5% to identify those who had received an mTBI from those who had not. In classification, sensitivity of 94.7% and specificity of 86.2%, AUROC score of 0.904

Physics Informed Machine Learning Improves Detection of HeadImpacts ⁴	mBTI (mild traumatic brain injury)	American football	PIML Convolutional Neural Network	6 variables	Negative predictive value and positive predictive values of 88 and 87% respectively (improved compared to traditional impact detectors on test datasets). Model reported best results to date for an impact detection algorithm for American football (F1 score of 0.95), could replace traditional video analysis for more efficient impact detection Peak network performance was at 100% additional synthetic data with NPV of 0.87 and PPV of 0.86
Machine Learning and Statistical Prediction of Pitching Arm Kinetics ²⁹	Throwing-related injuries	Baseball	Four supervised machine learning models (random forest, support vector machine [SVM] regression, gradient boosting machine, and artificial neural networks) were developed	Pitch velocity and 17 pitching mechanics	The gradient boosting machine (best performance): smallest RMSE (0.013%BWxH) and the most precise calibration (1.00 [95% CI, 0.999, 1.001]) The random forest model: largest RMSE (0.46%BWxH) and calibration (1.34 [95% 1.26, 1.42]). Regression model: The final model RMSE was 21.2%BW, calibration was 1.00 (95% CI, 0.88, 1.12), and was 0.51
Machine Learning for Predicting Lower Extremity Muscle Strain in National Basketball Association Athletes ³⁰	Lower Extremity Muscle Strain	Basketball	Random forest, extreme gradient boosting (XGBoost), neural network, support vector machines, elastic net penalized logistic regression, and generalized logistic regression.	Demographic characteristics, prior injury documentation and performance metrics	XGBoost models (higher AUC on internal validation data and a slightly higher Brier score): 0.840 (95% CI, 0.831-0.845), was best-performing algorithm (highest overall AUC, with decent calibration and Brier scores) Conventional logistic regression (significantly lower AUC on internal validation than Random forest and XGBoost): 0.818; (95% CI, 0.817-0.819) Calibration slope of models: from 0.997 for neural network to 1.003 for XGBoost (excellent estimation for all models) Brier score of models: from 0.029 for random forest to 0.31 for multiple models (excellent accuracy)
Forecasting football injuries by combining screening, monitoring and machine learning ³¹	Non-contact time-loss injuries	Football (soccer)	Gradient boosted model	Basic player information, screening, monitoring exposure, special vulnerability	Cross validated performance of gradient boosted model (ROC area under the curve 0.61) Holdout test set performance was similar (ROC area under the curve 0.62) shows its generalizability

predictors as well as performance. Abbreviations such as SIMS, RTP, CI, RMSE, SVM, QIC, vGRF, KFA, KEM, KPA, KJM, ROC refer to Soccer Injury Movement Screen, Return-to-Play, Confidence Interval, Root Mean Square Error, Support Vector Machine, Quasi-Information Criterion, vertical Ground Reaction Force, Knee Flexion Angle, Knee Extension Moment, Knee Power Absorption, Knee Joint Moment, Receiver Operating Characteristic respectively. To see full chart, see Supplementary Material]

The studies analyzed cover a wide range of injuries, including muscle, bone, non-contact, and contact. The most common injury type was related to muscle, specifically in the lower limb as most of the analyzed sports require the use of the athletes' lower limbs. These lower limb injuries included muscle, ligament, and bone injuries and different, specific areas, such as the knees and ankles. There were 20 studies (12-31) analyzing lower limb injuries, but within those studies, 7 (13, 15, 20, 22, [23], 25, 31) were assumed because the type of injury was unspecified. They were taken as such through the great use of lower limbs within the sport. The most common sports were football and basketball where almost half of the studies studied football and/or basketball. There were 5 articles (15, 17, 19, 26, 31) about only soccer, 2 articles (23, 30) about basketball only, and 2 (22, 28) articles analyzing soccer and basketball. There were appearances of different types of predictors such as performance tests, history of the patient injury-wise, baseline characteristics, athlete measurements and anthropometric attributes. Most of

the studies were using baseline characteristics, a history of past injuries, and specific measurements with respect to the sport to allow for performance or movement tracking. There were 16 (13, 15-17, 19-20, 22, 24, 26-29, 31-34) articles that used results from a performance test or screening as risk factors. Other than those, the next most used risk factor was related to their athletic body and environment. Most of the analyzed studies included around 100 participants and 2 studies (21, 30) had over 1000 participants. Furthermore, many of these studies were novel in predicting injuries, and therefore, had no original, existing baselines to compare against. Thus, 10 studies (14, 17-18, 22, 26, 25, 31, 9, 3, 6) did not have baselines.

In the bar covering 2015-2020 in Table 1, all 11 studies analyzed in the reference paper were included except for one that was published before 2015, one that was not PubMed anymore and one that did not show up under the same search criteria in PubMed. As for the ones that were in the reference paper but seemed to be published after March 2020, they were still included in the 2015-March 2020 timeframe for consistency. Furthermore, there was an article that was not included in the reference paper but was published between 2015-2020 and matched the criteria for this literature review, and thus was also included. The reason behind the stratification into pie charts is to showcase the progression of algorithms in numbers since the publishing date of the reference paper, especially, deep learning which is an algorithm that has been developed further in these past few years. Within the studies analyzed, there was ultimately

an increase in the use of ML methods for injury prediction in sports. The bars covering 2015-2020 and 2020-2023 illustrate the ML technology and the purpose of that is to understand changes in ML trends, more specifically, from the publishing date of the reference paper. Presented in the bar covering 2015-2020 in Figure 1 are the articles used in the reference article. There were 9 articles (12-19, 36) from 2015 - March 2020 and 16 articles (20-35) from March 2020 - 2023. Between 2015 and 2020, mostly random forest and support vector machine were used. These 2 machine learning techniques were generally the most applied ones between 2015 and 2020 with 6 studies (12, 14-16, 19, 36) in that category. The other notable difference was an increase in adoption of deep learning techniques. There were 2 studies (13, 17) that used deep learning, accounting for 22.2% of the articles between 2015 and March 2020. Between March 2020 and 2023, the number of uses of deep learning within studies rose to 7 studies (22, 24, 29-30, 33-35), taking up 43.8% of all the ones studied and published within that time range. The amount of Random Forest or Support Vector Machine utilization remained the same at 6 (20-21, 25, 27-28, 32). Moreover, there was one appearance of Regression-based algorithm use and one other ML algorithm used compared to 2015-2020. The “Other” ML algorithms were generally gradient boost ones and more specifically, XGBoost models.

Regarding training strategy, 24 studies included pre-processing techniques (12-22, 24-36) Some studies did not and instead, considered all factors and predictor candidates to make predictions. There were 17 studies (13-16, 18-21, 24-27, 30-32, 34, 36) that were set for cross validation, the most accepted method to effectively train ML methods. Moreover, 15 articles (12-13, 15-16, 19-21, 23, 25, 27-28, 30-33) used (what does AUC stand for) AUC to assess results. There were articles using decision trees and/or random forest on different scenarios and conditions, but in general, they gave modest results where the AUC, if provided, was usually around 0.5 and 0.6 (13, 16, 20, 22, 8, 27). There were some incidences for the models using decision trees where the value would reach 0.8 or above (17, 8). XGboost models, which are gradient boosting algorithms were also used on different conditions, but in general, provided high AUC scores of almost 1 (21, 30). Regarding studies using deep learning, almost all of them were analyzing different problems. Nonetheless, the AUC values of those studies were generally higher compared to other algorithms, but more so, after or starting in the later range of 2015-2020. In the earlier years of the range, the resulting AUC value was around 0.57 to 0.59 (14). After 2020, the results of the studies using deep learning (18, 23, 24, 30, 32, 33) presented reliability in their respective problems, higher AUC values of around 0.9 (3), calibration slopes of close to 1 (30), sensitivity about 95% and specificity of about 86% (33). In comparison to the XGBoost models, the objective AUC values of the deep learning models were lower. The articles with more than 1000 in sample size generally achieved

higher predictive ability. Variations in algorithm performance can be attributed to a difference in complexities or different sizes of data sets. Data set sizes were not kept constant due to some algorithms requiring specific sizes, such as deep learning needing a larger one to work effectively. The limitation of these studies is usually the lack of data available to work with. However, there are additional causes of complexity such as the sport’s nature, as there would not be the same predictive formats, intensity, and type of sport. The way the athlete obtains contact to injury also varied.

Discussion

Ultimately, the articles in this study showed that the field of sports injury prediction with machine learning algorithms is continuously growing and that AUC improvements are occurring as deep learning and XGBoost models are continued to be applied as well as developed. XGBoost especially is becoming a common algorithm used in sports injury prediction because the data samples do not have to be as large as those for deep learning algorithms. Furthermore, this algorithm has been proven to be successful through the combination of various decision trees. Comparing XGBoost to Random Forest, XGBoost has an advantage for unbalanced datasets, needs a lower number of initial hyperparameters, and generates a score called “Similarity score” (33). It is important because it shows the progression from the reference article which was published in March 2020 in terms of the increasing AUC values, the higher specificity and sensitivity values as well as the lower RMSE values. These findings matter because the occurrences of high accuracy represent ML’s potential consistent application within sports medicine and maintaining healthy practice. The appearance of moderate accuracy also offers importance because it provides room for improvement, but even more so, usefulness and disruptiveness depending on the problem and algorithm. This creates excitement and hope for the future of AI/ML algorithms being used widely in sports medicine. The occurrence of fatal athlete instances and impacts on parties can therefore be reduced or prevented now or in the near future. Society is also introduced to these algorithms.

The findings regarding the predictive ability of differing ML algorithms in this study are somewhat consistent with other literature reviews (2, 15, 29, 34, 32) including the reference paper. However, it is important to recognize inconsistency in their results of overall predictive ability within the field due to change in trends from less success in prediction to more success as deep learning and other methods of injury prediction advanced. Furthermore, inconsistency of results can also be due to different purposes of their investigation as well as their inclusion/exclusion criteria. Compared to the reference paper, this upward trend is maintained, but compared to other literature reviews, there is inconsistency as the field has not yet been delved into.

As the reference paper used only the PubMed database, this study also did to maintain consistency. However, that causes the search to be limited to the PubMed database and there may exist additional studies on the topic not included in this systematic review. Those articles could have created an impact on the overall result and judgment on not only ML, but also its clinical applicability. Some articles were missing information on sample size or results with a specific metric, consistent with the other articles although they matched the inclusion/exclusion criteria pushed by the reference paper. Another limitation is the lack of an established baseline in about half of the studies. This complicates the evaluation of the contribution of the different types of algorithms to performance improvement. Other limitations were the lack of information needed to form causal relationships to draw detailed and consistent information on ML for clinical application. There were many different scenarios within the studies analyzed such as different sports, training strategy, and sample sizes even when the machine learning algorithm used was the same. This creates difficulty in concluding the factor that should be responsible for differences such as if it should be the nature of the application itself or the particular ML method deployed. Similarly, the data set sizes also limit which ML algorithm is applied.

For future directions and recommendations, more detailed design labs are necessary for this investigation to further investigate causal relationships of achieving the best predictive results. The sensitivity could also be improved because although it is fairly high for some of the algorithms such as deep learning, an extremely high sensitivity is crucial for consistent clinical application. It is also important to realize that clinical application does not depend on the AUC, but on the nature of its application. Models with lower AUC values can still be useful in clinical practice because of that. The studies analyzed were not identical to one another, so the AUC results can have different implications. AUC was mainly used as the metric for the models due to its advantage in assessing imbalanced datasets as it takes into consideration performance over all possible classification thresholds (³⁵).

Conclusion

This study analyzed and extracted new information of the state-of-the-art ML algorithms and their improved uses within sports injury prediction. With consistent occurrences of higher AUC of up to a 0.2 or 0.3 increase and optimal metric values with newer algorithms, this progression from the reference article just 3 years ago provides insight and potential for the future. The study revealed that the number of articles has almost doubled with more of the portion attributed to deep learning. Nonetheless, more research and larger data sets to clearly test on are needed. The contribution of this work is to update the state-of-the-art of the field. This study is reproducible if the same criteria is used

because it is based off references and articles that are publicly available. Using similar experiments in the future with slight changes on these new algorithms has the potential to change the field in an upward direction.

Acknowledgements

The Lumiere program is acknowledged for providing the aid and guidance to conducting this study with thoroughness. The Lumiere mentor, Jorge Corral Acero, is acknowledged for his insightful comments which were crucial to achieving depth within the articles studied. The Lumiere writing coach, Rose Wang, is acknowledged for her detailed feedback on the format, vocabulary and grammar of this systematic review. The Lumiere publication strategist, Ayush Dhall, is acknowledged for his helpful information on the publication process. The Lumiere program manager, Srishti Bansal, is acknowledged for her fair and effective organization on the process of writing the research paper.

Supplemental Materials

Table 2 Complete ML Algorithms Prediction of Injuries Study Characteristics

Publishing Date	Article Title	Reference	KeyInfo	Injury	Sport	ML Technique/Model	Baseline	Size of Study/Sample Size	Training Strategy Predictor Selection	Predictors (Number, Type, ...)	Performance (Metrics, Performance vs Baseline)
2016	Predictors of Ulnar Collateral Ligament Reconstruction in Major League Baseball Pitchers [36]	DOI: 10.1177/0363546516643812	Purpose is to identify crucial predictors of UCL reconstruction	Ulnar collateral ligament injuries	Major League Baseball (MLB)	Binary logistic regression, naive Bayes, support vector machine, binary linear regression	Standard threshold of $-r_i$ 0.7 for strong collinearity Significant level of P.05	cohort of 104 pitchers	12 predictor variables entered into the logistic regression to train naive Bayes classifier and a linear support vector machine classifier. Machine learning model trained through 5-fold cross validation Predictor selection through data harvesting process from MLBAM website	14 predictor variables	The binary linear regression model: statistically significant (P 0.01), explained 19.9% of the variance in UCL reconstruction surgery and correctly classified 66.8% of cases Naive Bayes classification accuracy. 72% Support vector machine classification accuracy. 75%
2017	Importance of Training Load Measures in Injury Incidence of Professional Rugby League Athletes [12]	DOI 10.123 Ajapp 2016-0326	Purpose isto investigate ability of trainingload (TL) Lonitoring leasures to predictinjury	Time loss, soft tissue, overuse uries	Professional rugby (National Rugby League competition)	Generalized estimating equations (GEE) model and random forest	The area under the curve of the ROC where a value indicate 100% accuracy in the model predicting the target variable Model goodness of fit assesse by Quasi Likelihood under Independence Model Criterion (QIC) a	Twenty five professional rugby league players and 68 player seasons	training load data analyzed The data was partitioned into training validation and testing data sets (70/15/15%)	21 predictor variables different ones for different types of players	GEE Models: Adjustables OIC of 566.5 and p 0.001 to 0.091 Hit-up forwards QIC of 441.7 and p= 0.006 to 0.138 Outside backs - QIC of 406.6 end p=0.092 to 0.225. Wide-rurringforwards QIC of 410.5 and p 0.068 to 0.830 Random forest Mean (+SD) ROC for all the groups around 0.6 and 0.7.
2018	Predictive Modeling of Hamstring Strain Injuries in Elite Australian Footballers [13]	DOI 10.1249/MSS.0000000001527	Purpose of the study to investigate the effectiveness in predicting through using HSI (hamstring strain injury) risk factors in machine learning algorithms	Hamstring strain injuries	Australian Football	Supervised learning techniques (Naive Bayes Logistic regression, Random forest, Support vector machine Neural network)	N/A	2 cohorts, 186 (in 2013) and 176 (in 2015) Australian footballers	Demographic and injury history data collected Training is 10-fold cross validation, and no predictor selections	Eccentric hamstring strength, age, previous hamstring injury, between limb imbalance, previous ACL (anterior crucial ligament) injury, stature, mass, and primary playing position	-0.57 to 0.59 median AUC for any of the models

2018	Effective injury forecasting in soccer with GPS training data and machine learning [14]	DOI: 10.1371/journal.ne.0201264	Using GPS tracking technology, the information gathered is used to create an injury predictor.	Non-contact injuries	Soccer	Decision tree algorithm	Baseline B1 (assign as to an example while maintaining the distribution of classes), Baseline B2 (assigns non-injury class), Baseline B3 (assigns injury class). Baseline B4 (a classifier that assigned as 1 (injury) if $P(EWMA) \geq 0$, and 0 (no injury) if criteria not met)	26 Italian professional male players during 2013-2014 season	Monitored physical activity of players A training dataset T is made, Predictor selection by Recursive Feature Elimination with Cross-Validation in selecting TRAIN (30% of T), over sample to fix imbalance and find most relevant predictors Split TEST (70% of T) into 2 folds, f1 and f2 for stratified cross validation	55 features 18 daily features, 12 EWMA (Exponential Weighted Moving Average) features, 2 ACWR (Acute Chronic Workload Ratio) features, 12 MSWR (monotony) features, 1 previous injury feature	Decisiontree classifier DT has recall = 0.80+0.07 and precision = 0.50+0.11 on the injury dass, can predict 80% of injuries and can label a training session as an injury in 50% of the cases
2018	A Preventive Model for Muscle Injuries: A Novel Approach based on Learning Algorithms [15]	DOI: 10.1371/journal.ne.00001535	Purpose is to compare machine learning algorithms and select best performing one for identifying athletes at risk of lower extremity muscle injuries (MUSINJ).	Lower extremity muscle injuries (MUSINJ)	Professional soccer and handball 1	Decision tree algorithms (with Random tree as well), ensemble learning algorithms	AUC values (0.90-1.00), moderate (0.70-0.90), low(0.70-0.50) and fail (≤ 0.50)	Total of 132 male professional soccer and handball players	5-fold stratified cross validation (SCV technique) Screening evaluation for risk factors	52 features (Personal/individual risk factors, psychological risk factors, neuromuscular risk factors)	AD Tree achieved best performance in most of analyzed methods (AUC values of (0.6-0.7)
2019	A Preventive Model for Hamstring Injuries in Professional Soccer Learning Algorithms [16]	DOI: 10.1055/a-0826-1955	Purpose of the study is to compare the predictive ability of various machine learning techniques for identifying players at high risk of HSI's (Hamstring strain injury).	Hamstring strain injury	Professional soccer	Decision tree algorithms: J48, ADTree and SimpleCart.	N/A	96 male professional soccer players	Screening evaluation 3-fold stratified cross validation (SCV)	Modifiable and unmodifiable risk factors Personal risk factors, psychologicalrisk factors, neuromuscular risk factors	AUC of 0.837 for best performing model

2019	On-field player workload exposure and knee injury risk monitoring via deep learning [17]	DOI: 10.1016/j.jbiomech.2019.07.002	Using the CaffeNet convolutional neural network (CNN) model, multivariate regression of motion capture to 3D KJM (knee joint moments) for three sports movements were compared	Non-contact knee trauma	N/A	CaffeNet CNN (convolutional neural network)	N/A	Male and female athletes	Algorithm pretrained on ImageNet database. Sidestep movement 5 folds Rest of the movements: single 80:20-fold,	59 features	The strongest mean KJM correlation is for the left stance limb during idestepping (0.9179) which is also during sidestepping(= 0.8168)
2020	A Machine Learning Approach to Assess Injury Risk in Elite Youth Football Players [18]	DOI: 10.1249/MSS.0000000000002305	To assess injury risk based on measures with a machine learning model	All kind	Football (soccer)	Gradient boosting (XG-Boost)	Baseline characteristics of the players in form of means and standard deviations	734 male youth football players	Cross-validation Built model using our training data (random sample of 80% of all collected data). At the end, best-performing model wastested on our test data (remaining 20% of all collected data) Testing and questionnaires to conduct predictor selection	29 predictors from preseason test results	Metrics Precision, recall (sensitivity), accuracy (f1 score) Extreme gradient boosting model: Predict injury a precision of 84%, recall of 83% and anfl score of 83% in the training dataset. On the test data, the precision, recall and fl scores were all 85% (reasonable accuracy and sensitivity) Classifying injuries reasonably accurate in classifying injuries correctly
2020	Using machine learning to improve our understanding of injury risk and prediction in elite male youth football players [19]	DOI: 10.1016/j.jsams.2020.04.021	Compares logistic regression analysis with machine learning	Non-contact lower-limb injuries	Elite football	Decision trees: (J48con), an alternating decision tree (ADT) and a reduced error pruning tree (REP-Tree)	ZeroR classifier that obtained an AUC score of 0.494, specificity of 100% and sensitivity of 0%	355 elite youth football players (10-18 years old)	Pre-season neuromuscular screen To manage imbalance and skewed distributions, four resampling three classic ensemble, three bagging ensemble, three boosting ensemble and five cost-sensitive algorithms applied to data	Anthropometric measures of size, single leg countermovementjump (SLCMJ), single leghop for distance (SLHD), 75% hop distance and stick (75%Hop), Y-balance anterior reach and tuck jump assessment	Logistic regression on categorical data: For the univariate analysis, all variables reported an AUC 0.57, a sensitivity of 0% and a specificity of 100% With no collinearity, a multivariate analysis offered prediction with specificity of 94.5%, sensitivity of .1% and an AUC of 0.687. Best performing decisi tree model was with the bagging ensemble method on the J48 con decision tree base classifier. Cross validation gave an AUC of 0.663 (model correctly classified 74.2% of non-injured players and 55.6% of injured players)

2020 Machine Learning Outperform Logistic Regression Analysis to Predict Next-Season NHL Player Injury: Analysis of 2017 [20]	2020 Machine Learning Outperform Logistic Regression Analysis to Predict Next-Season NHL Player Injury: Analysis of 2017 [20]	Machine Learning Outperform Logistic Regression Analysis to Predict Next-Season NHL Player Injury: Analysis of 2017 [20]	DOI: 10.1177/23259671232596712	DOI: 10.1177/23259671232596712	Compares All logistic regression with machine learning of using post	All kind Hockey regression with machine learning of using post	kind Hockey forest, nearest neighbors, Naive Bayes, XGBoost, Top Three Ensemble good, ζ 0.9 excellent. AUC values compared using Tukey post hoc analysis	Hockey Random AUC forest, K nearest using a Riemann sum. position XGBoost, Top AUC values of 0.6-0.7 are 213 Naive Bayes, Riemann XGBoost, Top AUC values of 0.6-0.9 good, and factor ζ 0.9 ordinary-least-excellent. squares of AUC values regression compared using Tukey post hoc analysis	AUC calculated hockey using a trapezoidal 2109 Naive Bayes, Riemann sum. position XGBoost, Top AUC values of 0.6-0.7 are 213 Naive Bayes, Riemann sum. position XGBoost, Top AUC values of 0.6-0.9 good, and factor ζ 0.9 ordinary-least-excellent. squares of AUC values regression compared using Tukey post hoc analysis	2322 male hockey players trapezoidal 2109 All Riemann sum. position variables AUC values players and assessed for of 0.6-0.7 are 213 multicollinearity poor, 0.7-0.8 goalies using the fair, 0.8-0.9 riance inflation good, and factor ζ 0.9 ordinary-least-excellent. squares of AUC values regression compared context; all using Tukey variables with post hoc VIF of ζ 10 were analysis excluded k-fold cross-validation and ζ 0.9 excellent. AUC values compared using Tukey post hoc analysis	Injury data from database cohort 2109 All predictor features for position variables are goalie players and assessed for 213 multicollinearity Prior injury AUC of 0.937 using the count was the riance inflation most effective and factor (VIF) in an predictor in ordinary-least-future am aunt predicted next-season squares of injuries injury with an accuracy regression of 94.6% (SD, 0.5%, context; all and an accuracy of variables with ζ 10 were goalies analysis excluded k-fold cross-validation of 10 folds with each model. Split the data into 2 sets 90% as the training set and 10% as the test set. Using training set, accuracy, reliability, and responsiveness were tested Postural stability, strength, flexibility	35 features for non-goalie cohort and 14 for features for season are goalie cohort compared with for regression multicollinearity Prior injury AUC of 0.937 using the count was the riance inflation most effective and factor (VIF) in an predictor in XGB loost model ordinary-least-future am aunt predicted next-season squares of injuries injury with an accuracy regression of 94.6% (SD, 0.5%, context; all and an accuracy of variables with ζ 10 were goalies analysis excluded k-fold cross-validation of 10 folds with each model. Split the data into 2 sets 90% as the training set and 10% as the test set. Using training set, accuracy, reliability, and responsiveness were tested Postural stability, strength, flexibility	XGB oost had the highest AUC of 0.948 14 for predicting next season juryrisk, cohort compared with logistic regression which had an Prior injury AUC of 0.937 (P count was the ζ .0001). riance inflation most effective in an predictor in XGB loost model future am aunt predicted next-season of injuries injury with an accuracy of 94.6% (SD, 0.5%, all and an accuracy of variables with 96.7% (SD, 1.3% for VIF of ζ 10 were goalies excluded k-fold cross-validation of 10 with each Split the sets Initia validation strategy of test/train ROC AUC was 79.02%, validation cross
2020 Machine Learning Predict Lower Extremity Musculoskeletal Injury Risk in Student Athletes [21]	2020 Machine Learning Predict Lower Extremity Musculoskeletal Injury Risk in Student Athletes [21]	DOI: Learning to Predict pot.2020 signifi cant. Lower .576655 injury risk factors and Musculoskeletal model on Injury Risk student athletes. Athletes [21]	Identifying 10.3389/f most oskeleta n1 cant. 1 NCAA risk injur cant. 1 .576655 injury risk factors and developing a all, model on Men's Risk student football athletes. Soccer and	Muscul oskeleta n1 cant. 1 NCAA risk injur cant. 1 .576655 injury risk factors and developing a all, model on Men's Risk student football athletes. Soccer and Women 's V olleyb all	Divisio oskeleta n1 forest 1 NCAA injuries sports Basketball all, Men's football Soccer and Women 's V olleyb all	Random forest NCAA sports Basketball all, Men's football Soccer and Women 's V olleyb neural	N/A NCAA athletes females, 71 males Relative error in the that	Division I NCAA athletes females, 71 Test/train males validation strate; rwith 80:20 split (separating on subjects), hyperparameters tuned by GridSearch women's basketball	122 college Division I NCAA assessments athletes 51 females, 71 Test/train males validation strate; rwith 80:20 split (separating on subjects), hyperparameters tuned by GridSearch women's basketball	responsiveness were tested Postural stability, strength, I flexibility assessments metrics athletes 51 females, 71 Test/train males validation strate; rwith 80:20 split (separating on subjects), hyperparameters tuned by GridSearch women's basketball	The data contains 50 physical split metrics accuracy spanning secondary split strength, with k-fldd validation postural vali dation of an average stability, and ROC AUC of 80:20 split flexibility (separating on (dominant and subjects), non-dominant hyperparameters leg) combined with previous injury binary classification and demographic data Improved unequal model	Initia validation strategy of test/train split ROC AUC accuracy was 79.02%, secondary validation strength, with k-fldd cross postural vali dation of an average stability, and ROC AUC of 6890% flexibility (dominant and non-dominant leg) combined with previous binary classification Predicted results of most basketball injuries close to the actual so the network's prediction reliable.	
2021 Basketball Sports Injury Prediction Model Based on the Grey Theory Neural Network [22]	2021 Basketball Sports Injury Prediction Model Based on the Grey Theory Neural Network [22]	DOI: 10.1155/2021/1653 Prediction 093 network Based on model and the Grey odel Neural improves Network predictive ability	Combining Ankle 10.1155/2021/1653 network 093 injuries mapping model and the coupling odel improves predictive ability	Ankle and 's knee basket network injuries all mapping model and the coupling odel improves predictive	Women Grey 's network basket injuries all the	Grey neural network (RE) basket sense all smaller RE, the prediction is to the actual value (expressed in percentage)	Relative error 4 (RE) in the sense that the teams smaller the in total) RE, the closer the prediction is to the actual value (expressed in percentage)	4 women's basket the teams (45 the in total) methods the closer postevent prediction is to the actual based on data value already occurred (expressed in while percentage) prediction is what might, has yet to	Post event prediction and (45 preprediction test interval total) methods is used in postevent prediction prediction is optimizing based on data model in gray already occurred theory. It is a while good predictor preprediction is a of sports prediction on injuries, but is what might, but not for the has yet to occur average number injuries	demographic data Improved unequal test interval model are is used in results, prediction by injury is optimizing the on gray is already occurred theory. It is a while good predictor preprediction is a of sports on injuries, but is what might, but not for the occur average number of injuries	Predicted results of most basketball injuries are close to the actual results, so the network's by injury prediction based optimizing the on gray is reliable. model in gray theory. It is a good predictor of sports injuries, but is for the of		

2021	Injury Prediction in Competitive Runners With Machine Learning ²⁴	DOI: 10.1123/ajp.2020-0518	Using machine learning to predict injuries in runners, based on training logs.	Unspecific running-related injuries	Competitive running	Extreme Gradient Boosting, or XGBoost	An AUC of 1.0 is a perfect prediction mode, 0.5 indicates random guessing. The closer the AUC is to 1, the better the prediction mode, AUC scores of 0.7 and higher indicate high or strong significance in the field of sports sciences.	Detailed training log of 77 runners (27 women and 50 men)	Bagging approach. Then, prediction is calculated as the mean of all predictions in participating models to get a wider representation of healthy examples than with a single model	Objective data from global positioning system watch (eg, duration, distance) and subjective data about the exertion and success of the training. Second, a training week was summarized by 22 aggregate features, and a time window of 3 weeks before the injury was considered	Average AUC scores of 0.729 and 0.724 for validation and test sets of day approach, and AUCs of 0.783 and 0.678 for the validation and test sets of the week approach
2021	Combining Inertial Sensors and Machine Learning to Predict vGRF and Knee Biomechanics during a Double Limb Jump Landing Task ²⁴	DOI: 10.3390/s21134383	Aims to develop multi-sensor machine learning algorithms for prediction	Anterior cruciate ligament	N/A	Support vector machines (SVMs), artificial neural networks (ANNs), and generalized linear models	If the clinical difference was larger than the error, there is more confidence in the algorithm. Otherwise, the algorithm may not work accurately enough. Benchmark of 0.8 for high accuracy of R ²	Twenty six healthy college students (25 female)	Participants performed jump landing task for data collection. Processed all IMU data and model trained with custom MATLAB scripts. Performed k-fold cross-validation (n = 10) to each multi-feature model selected (randomly assigned each limb-trial to 1 of 10 folds), then, features were extracted within the ROI from every high-g accelerometer and gyroscope time-series (x axis, axis, Z axis, and the resultant). A full list of the 14 features	Max, time to max, max prominence, width of max, min, time to min, min prominence, width of min, max-min difference, max-min times difference, start value, stop value, standard deviation, area under the curve	Across vGRF, KFA, KEM, and KPA, both multiple feature models had an RMSE smaller than the clinical difference, creating confidence in their use. Both single feature models had an RMSE larger than the clinical difference, limiting their use. Although the R ² for the KEM and KPA models did not exceed the 0.8 for high accuracy, the error also did not exceed the expected differences between ACLR and healthy control limbs, thus the models still have clinical value.

2021	New Machine Learning Approach for Detection of Injury Risk Factors in Young Team Sport Athletes ¹⁶	DOI: 10.1055/a-1231-5304	The purpose is to show how predictive machine learning methods can detect sport injury risk factors in a data-driven approach	Lower extremity injuries	Basketball and floorball	random forest and L1 regularized logistic regression	AUC-ROC of 1.0 for perfect prediction and 0.5 for purely random prediction	162 females and 152 males	10-fold cross-validation for training data, normalization and imputation for each fold and for test data, normalization K-nearest neighbor imputation where the k value was 10	For random forest, 12 consistent injury predictors For L1-regularized logistic regression, 20 consistent injury predictors	For random forest, the mean AUC-ROC value was 0.63 and 0.94 for the training data (values of real response higher than the randomized mean AUC-ROC 0.48) confirms significance For logistic regression, the mean AUC-ROC value was 0.65 and 0.76 for the training data (values of real responses higher than the randomized ones, mean AUC-ROC 0.50)
2022	Predictive Modeling of Injury Risk Based on Body Composition and Selected Physical Fitness Tests for Elite Football Players ²⁶	DOI: 10.3390/cm1164923	Study discusses on the topic of regression rather than classifier, so the question is framed towards the amount of injuries and not whether injuries will occur or not Linear, so might not be able to capture complexity that can be done using non-linear (for example: neural networks are good at this)	All kind	Football (soccer)	Linear regression models (Classic regression models (OLS), shrinkage regression, step wise regression, LASSO)	N/A	Physical fitness tests 36 players of professional football team	Physical tests The taxiing strategy is leave one-out cross validation	22 independent variables such as players' information, body composition, physical fitness, and one dependent variable, the number of injuries per season	For elastic net regression, its prediction error was small (RMSE = 0.633) and the amount of predictors was decreased due to the method S characteristics The use of shrinkage models (Ridge, LASSO, and elastic net) caused quick decrease in error Of (an improvement in the model's predictive ability) The Ridge model was the best-performing model for predicting injuries (RMSE error was 0.698)
2022	Predicting ACL Injury Using Machine Learning Data From an Extensive Screening Test Battery of 880 Female Elite Athletes ³³	DOI: 10.1177/03635465221112095	Aims to predictability of machine learning algorithm on a large set of risk factor data for anterior cruciate ligament (ACL) injury	Arterial cruciate ligament	Handball and softball	Random forest, L2-regularized logistic regression, and support vector machines (SVMs) with linear and nonlinear kernel	AUC-ROC (used with imbalanced class distributions): Excellent (0.90-1), good (0.80-0.89), fair (0.70-0.79), poor (0.60-0.69), or fail (0.50-0.59)	451 soccer and 429 handball players	Screening tests 5-fold cross validation	Player baseline characteristics elite playing experience, history of any previous injuries and measurements of anthropometric CS, strength, flexibility, and balance.	Linear SVM (without any imbalance handling) obtained an AUC-ROC value of 0.63 (highest mean) With all 4 classifier great differences between minimum and maximum AUC-ROC values during repetitions because of random CI loss-validation splits. Training AUC-ROC values were very high for Random forest and SVMs, but with logistic regression, control of overfitting was better.

2022	Detecting Injury Risk Factors with Algorithmic Models in Elite Women's Pathway Cricket ²⁷	DOI:10.1055/1502-6824	Purpose is to explore the ability of algorithmic models to identify important risk factors that may not have been realized otherwise.	N/A	Cricket	Decision tree and random forest	The higher the AUC, being between 0 and 1, the better the predictive ability 0.5 indicates that the prediction is pure chance and 1 indicates perfect prediction.	17 players on the England and Wales Cricket Board (ECB) women's international development pathway	Daily load data collected For model parameter optimization, ten-fold cross validation was applied on randomly selected training data (70% of the total) and for model validation, remaining data (30% of total) was used.	Traditional algorithm decision tree: a minimum of 20 splits and 7 variables all owed in any leaf, with a maximum depth of 30, including 1064 observations from 47 input variables Best-performing random forest model: 100 trees with 8 variables tried at each split and included 1,064 observations (null values were excluded) from 47 input variables.	Decision tree: poor overall probability of accurately predicting injury with the training data (56 % for each rule). On the testing data set (30% of the data randomly split), the conditional algorithm performed poorly (AUC of 0.66, but still slightly better than the traditional algorithm (AUC of 0.57)). Best performing random forest model: On the testing data set, conditional algorithm (AUC of 0.72) performed poorly, but still slightly better than the traditional algorithm (0.65)
2022	Impact of Gender and Feature Set on Machine-Learning-Based Prediction of Lower-Limb Overuse Injuries Using a Single Trunk-Mounted Accelerometer ²⁸	DOI: 10.3390/s22082860	Predicting lower-limb overuse injury (LLOI) using a machine learning model.	Lower-limb overuse injury (LLOI)	Dance, track and field, gymnastics, swimming, basketball, handball, soccer, and volleyball	Logistics regression, support vector, trees	AUC of 0.5 represents random guessing	204 first-year undergraduate students (141 males, 63 females) from two academic years (2019-2020 and 2020-2021)	Data collection through Cooper test Min/max normalization for feature selection Training Model we retrained on the entire dataset or a gender-specific subset of it. A six-fold CV (with a 3-fold internal CV) was implemented and the option of 30 PCA components was omitted for the female-specific models (smaller amount of female data compared to male or mixed-gender).	Basic characteristics (weight, height, gender, previous injuries, and whether they wore insoles) and tri-axial acceleration measurements over the L3 to L5 spinal segments while performing the running Cooper test.	Logistic regression (LR) model was best-performing in terms of AUC score (mean AUC score of 0.645+0.056). (used entire set of features), mean Brier score of 0.190+0.021. Second highest is for the models using statistical features, lowest-performing model in terms of AUC only using sports-specific features. Support vector machine models lowest mean Brier score (best results) for the female-specific, male-specific, and all-data models no matter the feature type or amount

2022	A machine learning approach to identify risk factors for running related injuries study protocol for a prospective longitudinal cohort trial ²⁵	DOI: 10.1186/s13102-022-00426-0	Machine learning approach used to analyze biomechanical, biological, and loading parameters for identification of risk factors and patterns	Unspecific running-related injuries	Competitive running	Deep Gaussian Covariance Network (DGCN)	N/A	Female and male runners aged 18 years and older with a minimum weekly training volume of 20 km	Performance tests and questionnaires Batch training	Internal (e.g., anatomy, biomechanics, musculoskeletal tissue quality) and external characteristics (e.g., environment, surface, footwear)	Deep Gaussian Covariance Network (DGCN)
2022	A deep learning based approach to agnose mild traumatic brain injury using audio classification ³²	DOI: 10.1371/journal.pone.0274395	Proposes the extraction of Mel Frequency Cepstral Coefficient (MFCC) features from audio recordings of the speech of athletes engaging in rugby union and diagnosed with an mTBI Of not.	mBTI (mild traumatic brain injury)	Rugby	Bidirectional long short-term memory attention (Bi-LSTMA) deep learning model	N/A	46 athletes from a university rugby team	Neurological screening MFCC features split at participant-level into train (60%), validation (20%), and test (20%) sets. Bi-LSTM-A trained with the PSO algorithm to optimize the model hyper-parameters	38 different vocal features were investigated to assess if they indicate the existence of a brain injury.	Little-to-no overfitting in the training process, so the training process on observed data would allow the model to generalize well for classifying unseen data (good reliability) Overall accuracy of 89.5% to identify those who had received an mTBI from those who had not. In classification, sensitivity of 94.7% and specificity of 62% AUROC score of 0.904.
2022	Physics-Informed Machine Learning Improves Detection of Head Impacts ²²	DOI: 10.1007/s10439-022-02911-6	By simulating head impacts numerically with a head-neck model, synthetic impacts can be created on the impact data from mouthguards	mBTI (mild traumatic brain injury)	American football	PIML, Convolutional Neural Network	N/A	Data from 12 collegiate players, 49 high school players, both over 17 practice and game days	Field data was separate into training validation, and testing (70-15-15% split). Synthetic data was not used for validation or testing to prevent interference with the algorithm's improvement in detecting real-world cases. Augmentation approach to improve the balance of true and false positive samples. Dataset consisted of comma separated value (csv) files that contain the signals recorded mouth-guard kinematic data	6 variables: the X, y, Z components of line an acceleration, three components of angular acceleration	Negative predictive value and positive predictive values of 88 and 87% respectively (improved compared to traditional impact detectors n test datasets). Model reported best results to date for an impact detection algorithm for American football (F1 score of 0.95), could replace traditional video analysis for more efficient impact detection Peak network performance was at 100% additional synthetic data with NPV of 0.87 and PPV of 0.86

2022	Machine Learning and Statistical Prediction of Pitching Arm Kinetics ³⁰	DOI: 10.1177/03635466187711054506	Aims to identify 5 variables impact elbow valgus torque and shoulder distraction force the most.	Throwing-related injuries	Baseball	Four supervised machine learning models (random forest, support vector machine [SVM] regression, gradient boosting machine, and artificial neural networks) were developed	N/A	A total of 168 pitchers (21% were left-handed, 80% were in high school)	Biomechanical evaluation on athletes. All machine learning models, except artificial neural networks: internally validated through 10-fold cross-validation. Artificial neural networks: internally validated through 100 replications.	All models used predictor variables: pitch velocity and 17 pitching mechanics	The gradient boosting machine (best performance): smallest RMSE (0.013% BWXH) and the most precise calibration (1.00 [95% CI, 0.999, 1.001]) The random forest model: largest RMSE (0.46% BWXH) and calibration (1.34 [95% CI, 1.26, 1.42]). Regression model: The final model RMSE was 21.2% BW, calibration was 1.00 (95%CI, 0.88, 1.12), and r2 was 0.51
2022	Machine Learning for Predicting Lower Extremity Muscle Strain in National Basketball Athletes ³⁰	DOI: 10.1177/23259671221111742	Aims to characterize the epidemiology of time-loss lower extremity muscle strains (LEMSs) and explore the possibility of a machine-learning model in predicting injury risk.	Lower Extremity Muscle Strain	Basketball	Random forest, extreme gradient boosting (XGBoost), neural network, support vector machines, elastic net penalized logistic regression, and generalized logistic regression	An AUC of 0.70 to 0.80 is acceptable, an AUC of 0.80 to 0.90 is excellent.	2103 NBA athletes	Data from online platforms. Models trained with 10-fold cross-validation repeated 3 times. Recursive feature elimination (RFE) using a random forest algorithm to select the most relevant features and eliminating variables with high collinearity within high-dimensional data	Demographic characteristics (age, career length, and player position), prior injury documentation (recent and remote injury history), and performance metrics (3-point attempt, free throw attempt rate, etc)	XGBoost models (higher AUC on internal validation data and a slightly higher Brier score): 0.840 (95% CI, 0.831-0.845), was best-performing algorithm (highest overall AUC, with decent calibration and Brier scores) Conventional logistic regression (significantly lower AUC on internal validation than Random forest and XGBoost): 0.818; (95%CL 0.817-0.819) Calibration slope of models: from 0.997 for neural network to 1.003 for XGBoost (excellent estimation for all models). Brier score of models: from 0.029 for random forest to 0.31 for multiple models (excellent accuracy)

References

- 1 Gough, *Global sports market revenue 2028*, 2024, <https://www.statista.com/statistics/370560/worldwide-sports-market-revenue/>.
- 2 J. Lemoyne, C. Poulin, N. Richer and A. Bussi eres, *The Journal of the Canadian Chiropractic Association*, 2017, **61**, 88–95.
- 3 A. Guermazi, F. W. Roemer, P. Robinson, J. L. Tol, R. R. Regatte and M. D. Crema, *Radiology*, 2017.
- 4 *What Are The Long-Term Effects Of Sports Injuries? - South Carolina Sports Medicine and Orthopaedic Center - Charleston, SC*, 2023, <https://scsportsmedicine.com/blog/what-are-the-long-term-effects-of-sports-injuries>.
- 5 M. Putukian, *Mind, Body and Sport: How being injured affects mental health*, 2014, <https://www.ncaa.org/sports/2014/11/5/mind-body-and-sport-how-being-injured-affects-mental-health.aspx>.
- 6 S. Weidman, *The 4 Deep Learning Breakthroughs You Should Know About*, 2017, <https://towardsdatascience.com/the-5-deep-learning-breakthroughs-you-should-know-about-df27674ccdf2>.
- 7 H. Van Eetvelde, L. D. Mendon a, C. Ley, R. Seil and T. Tischler, *Journal of Experimental Orthopaedics*, 2021, **8**, 27.
- 8 *Advantages and Disadvantages of Logistic Regression*, 2020, <https://www.geeksforgeeks.org/advantages-and-disadvantages-of-logistic-regression/>.
- 9 D. Niklas, *Random Forest: A Complete Guide for Machine Learning | Built In*, <https://builtin.com/data-science/random-forest-algorithm>.
- 10 D. K., *Top 4 advantages and disadvantages of Support Vector Machine or SVM*, 2023, <https://dhirajkumarblog.medium.com/top-4-advantages-and-disadvantages-of-support-vector-machine-or-svm-a3c06a2b107>.
- 11 B. Sushman, *Advantages of Deep Learning, Plus Use Cases and Examples | Width.ai*, 2021, <https://www.width.ai/post/advantages-of-deep-learning>.
- 12 D. Whiteside, D. N. Martini, A. S. Lepley, R. F. Zernicke and G. C. Goulet, *The American Journal of Sports Medicine*, 2016, **44**, 2202–2209.
- 13 H. R. Thornton, J. A. Delaney, G. M. Duthie and B. J. Dascombe, *International Journal of Sports Physiology and Performance*, 2017, **12**, 819–824.
- 14 J. D. Ruddy, A. J. Shield, N. Maniar, M. D. Williams, S. Duhig, R. G. Timmins, J. Hickey, M. N. Bourne and D. A. Opar, *Medicine and Science in Sports and Exercise*, 2018, **50**, 906–914.
- 15 A. Rossi, L. Pappalardo, P. Cintia, F. M. Iaia, J. Fern andez and D. Medina, *PLOS ONE*, 2018, **13**, e0201264.
- 16 A. L opez-Valenciano, F. Ayala, J. M. Puerta, M. B. A. DE Ste Croix, F. J. Vera-Garc a, S. Hern andez-S anchez, I. Ruiz-P erez and G. D. Myer, *Medicine and Science in Sports and Exercise*, 2018, **50**, 915–927.
- 17 F. Ayala, A. L opez-Valenciano, J. A. G amez Mart ın, M. De Ste Croix, F. J. Vera-Garc a, M. D. P. Garc a-Vaquero, I. Ruiz-P erez and G. D. Myer, *International Journal of Sports Medicine*, 2019, **40**, 344–353.
- 18 W. R. Johnson, A. Mian, D. G. Lloyd and J. A. Alderson, *Journal of Biomechanics*, 2019, **93**, 185–193.
- 19 N. Rommers, R. R ossler, E. Verhagen, F. Vandecasteele, S. Verstockt, R. Vaeyens, M. Lenoir, E. D’Hondt and E. Witvrouw, *Medicine and Science in Sports and Exercise*, 2020, **52**, 1745–1751.
- 20 J. L. Oliver, F. Ayala, M. B. A. De Ste Croix, R. S. Lloyd, G. D. Myer and P. J. Read, *Journal of Science and Medicine in Sport*, 2020, **23**, 1044–1048.
- 21 B. C. Luu, A. L. Wright, H. S. Haerberle, J. M. Karnuta, M. S. Schickendantz, E. C. Makhni, B. U. Nwachukwu, R. J. Williams and P. N. Ramkumar, *Orthopaedic Journal of Sports Medicine*, 2020, **8**, 2325967120953404.
- 22 M. Henriquez, J. Sumner, M. Faherty, T. Sell and B. Bent, *Frontiers in Sports and Active Living*, 2020, **2**, 576655.
- 23 F. Zhang, Y. Huang and W. Ren, *Journal of Healthcare Engineering*, 2021, **2021**, 1653093.
- 24 C. Cr, B. Nt, A.-L. C. K. Aw, M. Mj and P. Da, *Sensors (Basel, Switzerland)*, 2021, **21**, year.
- 25 A. L. Rahlf, T. Hoenig, J. St urznickel, K. Cremans, D. Fohrmann, A. Sanchez-Alvarado, T. Rolvien and K. Hollander, *BMC sports science, medicine & rehabilitation*, 2022, **14**, 75.
- 26 F. Martins, K. Przednowek, C. Fran a, H. Lopes, M. de Maio Nascimento, H. Sarmento, A. Marques, A. Ihle, R. Henriques and R. Gouveia, *Journal of Clinical Medicine*, 2022, **11**, 4923.
- 27 L. Goggins, A. Warren, D. Osguthorpe, N. Peirce, T. Wedatilake, C. McKay, K. A. Stokes and S. Williams, *International Journal of Sports Medicine*, 2022, **43**, 344–349.
- 28 S. Bogaert, J. Davis, S. Van Rossom and B. Vanwanseele, *Sensors (Basel, Switzerland)*, 2022, **22**, 2860.
- 29 B. Gs, M. J, H. T, N. Kf, R. Rd and C. Gs, *Sports medicine (Auckland, N.Z.)*, 2022, **52**, year.
- 30 Y. Lu, A. Pareek, O. Z. Lavoie-Gagne, E. M. Forlenza, B. H. Patel, A. K. Reinholz, B. Forsythe and C. L. Camp, *Orthopaedic Journal of Sports Medicine*, 2022, **10**, 2325967122111742.
- 31 A. Hecksteden, G. P. Schmartz, Y. Egyptien, K. Aus der F nten, A. Keller and T. Meyer, *Science & Medicine in Football*, 2023, **7**, 214–228.
- 32 W. Schmid, Y. Fan, T. Chi, E. Golanov, A. S. Regnier-Golanov, R. J. Austerman, K. Podell, P. Cherukuri, T. Bentley, C. T. Steele, S. Schodrof, B. Aazhang and G. W. Britz, *Journal of Neural Engineering*, 2021, **18**, year.
- 33 Gupta, *XGBoost versus Random Forest | Qwak*, 2021, <https://www.qwak.com/post/xgboost-versus-random-forest>.
- 34 H. Compton, J. Delaney, G. Duthie and B. Dascombe, *International Journal of Sports Physiology and Performance*, 2016, **12**, year.
- 35 *F1 Score vs ROC AUC vs Accuracy vs PR AUC: Which Evaluation Metric Should You Choose?*, <https://neptune.ai/blog/f1-score-accuracy-roc-auc-pr-auc>.