

Non-contact lower limb injuries in Rugby Union: A two-year pattern recognition analysis of injury risk factors

Evans, Seren; Owen, Robin; Whittaker, Gareth; Davis, Oran Elphinstone; Jones, Eleri; Hardy, James; Owen, Julian

PLoS ONE

DOI:
[10.1371/journal.pone.0307287](https://doi.org/10.1371/journal.pone.0307287)

Published: 24/10/2024

Publisher's PDF, also known as Version of record

[Cyswllt i'r cyhoeddiad / Link to publication](#)

Dyfyniad o'r fersiwn a gyhoeddwyd / Citation for published version (APA):
Evans, S., Owen, R., Whittaker, G., Davis, O. E., Jones, E., Hardy, J., & Owen, J. (2024). Non-contact lower limb injuries in Rugby Union: A two-year pattern recognition analysis of injury risk factors. *PLoS ONE*, 19(10), Article e0307287. <https://doi.org/10.1371/journal.pone.0307287>

Hawliau Cyffredinol / General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

RESEARCH ARTICLE

Non-contact lower limb injuries in Rugby Union: A two-year pattern recognition analysis of injury risk factors

Seren Lois Evans^{1*}, Robin Owen², Gareth Whittaker³, Oran Elphinstone Davis³, Eleri Sian Jones⁴, James Hardy⁴, Julian Owen¹

1 Institute for Applied Human Physiology, School of Human and Behavioural Sciences, Bangor University, Bangor, United Kingdom, **2** School of Health and Sport Sciences, Liverpool Hope University, Liverpool, United Kingdom, **3** Welsh Rugby Union, Colwyn Bay, United Kingdom, **4** Institute for Psychology of Elite Performance, School of Human and Behavioural Sciences, Bangor University, Bangor, United Kingdom

* seren.evans@bangor.ac.uk



Abstract

The cause of sport injuries are multifactorial and necessitate sophisticated statistical approaches for accurate identification of risk factors predisposing athletes to injury. Pattern recognition analyses have been adopted across sporting disciplines due to their ability to account for repeated measures and non-linear interactions of datasets, however there are limited examples of their use in injury risk prediction. This study incorporated two-years of rigorous monitoring of athletes with 1740 individual weekly data points across domains of training load, performance testing, musculoskeletal screening, and injury history parameters, to be one of the first to employ a pattern recognition approach to predict the risk factors of specific non-contact lower limb injuries in Rugby Union. Predictive models (injured vs. non-injured) were generated for non-contact lower limb, non-contact ankle, and severe non-contact injuries using Bayesian pattern recognition from a pool of 36 Senior Academy Rugby Union athletes. Predictors for non-contact lower limb injuries included dorsiflexion angle, adductor strength, and previous injury history (area under the receiver operating characteristic (ROC) = 0.70) Dorsiflexion angle parameters were also predictive of non-contact ankle injuries, along with slower sprint times, greater body mass, previous concussion, and previous ankle injury (ROC = 0.76). Predictors of severe non-contact lower limb injuries included greater differences in mean training load, slower sprint times, reduced hamstring and adductor strength, reduced dorsiflexion angle, greater perceived muscle soreness, and playing as a forward (ROC = 0.72). The identification of specific injury risk factors and useable thresholds for non-contact injury risk detection in sport holds great potential for coaches and medical staff to modify training prescriptions and inform injury prevention strategies, ultimately increasing player availability, a key indicator of team success.

OPEN ACCESS

Citation: Evans SL, Owen R, Whittaker G, Davis OE, Jones ES, Hardy J, et al. (2024) Non-contact lower limb injuries in Rugby Union: A two-year pattern recognition analysis of injury risk factors. *PLoS ONE* 19(10): e0307287. <https://doi.org/10.1371/journal.pone.0307287>

Editor: Holakoo Mohsenifar, Iran University of Medical Sciences, ISLAMIC REPUBLIC OF IRAN

Received: December 22, 2023

Accepted: July 3, 2024

Published: October 24, 2024

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pone.0307287>

Copyright: © 2024 Evans et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its [Supporting information](#) files.

Funding: The authors received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

Introduction

The lower limb has the highest incidence of injury in Rugby Union [1–4], resulting in 50% of all days lost due to injury in international Rugby Union [5]. It is well established that as the burden of injury increases there are significant reductions in team performance [6]. Therefore, being able to identify when the risk of certain injuries are higher through modes of athlete monitoring could be invaluable in not only reducing the frequency of these injuries, but to inform injury prevention strategies and increase player availability, subsequently impacting team success [7]. Inevitably, the prediction of injuries has been investigated by applied sport scientists, who have presented a priori of reasons for collecting data on specific factors and utilizing linear based analyses [8]; undoubtedly this activity has been important in laying the foundations of existent sport injury literature. However, the explanatory power of these analyses may not infer the predictive power needed for sport injury risk detection [8–11]. As sport injuries are evidently complex and multifaceted [12], examining isolated risk factors associated with sport injury may be a reductionistic approach and unlikely to direct optimal injury prevention strategies [13, 14]. Despite the established evidence pointing towards the complexity of sport injury, there is limited research adopting a longitudinal, multifactorial approach in identifying predictive variables for injury whilst acknowledging the likelihood of non-linear interactions of variables, particularly with reference to specific anatomical locations. Deciphering which combination of variables contribute to increased injury risk by specific location or diagnosis (i.e. ankle injury, hamstring injury, non-contact lower limb injury), may aid in facilitating a more targeted approach to injury prevention.

Adopting a targeted approach to forecasting injuries means incorporating established and potential injury risk factors to accurately identify the combination of predictive variables for a specific injury. Performance based attributes such as greater maximal sprinting speed and aerobic fitness testing have been found to reduce the risk of lower limb injuries in sport [15, 16]. Furthermore, data derived from musculoskeletal screening have been identified as injury risk factors of the lower limb; for example ankle dorsiflexion angle deficits, and reduced adductor strength have previously been associated with an increased risk of lower extremity injuries such as patellar tendinopathy and adductor strains [17–20]. Monitoring athlete perceived ratings of muscle soreness is also an indicator of player training status and combined with factors of fatigue and musculoskeletal screening have the potential to identify greater likelihood of lower limb injury [21]. Some evidence also suggests that decreased eccentric hamstring strength is associated with greater risk of hamstring injury [22–24], with a well-established link between previous injury history and re-injury [25–29]. Despite the important role of linear analyses within early research in establishing injury risk factors, they may only demonstrate association between the variables and injury, and not their predictive importance [8, 10]. Therefore, a targeted pattern recognition approach to data handling in injury risk related research is pivotal for constructing robust injury modelling frameworks; allowing for the analysis of larger datasets including interactions and non-linear relationships of established modifiable and non-modifiable injury risk factors to formulate specific, interpretable injury risk profiles [8, 30].

Pattern recognition is a contemporary statistical approach whereby algorithms enable modelling of complex, multiple non-linear interactions between variables [31, 32]. These data handling methods overcome the limitations of typical logistic regression analyses previously used in “predicting” injuries in team sports, and further facilitate the development of rigorous, generalisable injury risk models inclusive of predictive variables. Emerging evidence highlight the efficacy of pattern recognition analyses in discerning predictive variables associated with heightened injury risk across various team sports [10, 32, 33]. Being able to effectively utilise

absolute and weekly change values of athlete monitoring components to communicate with stake holders which athletes are at an increased risk of injury could be invaluable for both coach education and athlete welfare. To our knowledge, there are no studies utilising a pattern recognition approach to identify lower limb non-contact injury risk factors in Rugby Union. Therefore, the aim of this study was to investigate whether a pattern recognition approach could be used to identify in-depth predictive variables that increase the risk of lower limb non-contact injuries in Rugby Union across multiple seasons based on data spanning training load parameters, performance testing results, previous injury history, and musculoskeletal screening. Pattern recognition algorithms will be used to create models comprising of the most predictive set of risk factors for lower limb non-contact injury, non-contact hip/groin injuries, non-contact hamstring injuries, non-contact knee injuries, non-contact ankle injuries and severe lower limb non-contact injuries.

Methods

Data from thirty-six male regional Rugby Union players, representing the Senior Academy, were collected across two consecutive seasons (2018–2019, 2019–2020). Players attended three full days of training which included unit skills, speed and gym sessions lead by strength and conditioning and performance coaches, and two evening rugby-based sessions. League matches were played at the weekends from September until mid-May with full-time Physiotherapy provision and match-day doctors. Semi-professional Rugby Union is often seen as a key stepping-stone into professional sport and for player development, with players often being paid and some may train full-time whilst others will have primary full-time employment elsewhere [34]. Inclusion criteria for this study required all athletes to be in full-time training to ensure consistency across data points collected and to reduce the likelihood of missing data. All athletes were male semi-professional Rugby Union players and had to be over the age of 18. Each player provided written informed consent for the collection of: baseline anthropometric and performance testing; retrospective and prospective injury data; musculoskeletal screening; and training load parameters. For each individual player, 1740 in-season weekly data points were collected. Late season data was omitted due to there being numerous missing data points amidst the COVID-19 pandemic in the second season of this study. Player data was excluded if they had multiple missing data points across the season within measures training load, performance testing, musculoskeletal screening and injury history parameters. Part-time athletes were excluded from this study, and any athletes who left the team during the data collection period were also excluded. The study was approved by the Bangor University Ethics Committee of the School of Sport, Health and Exercise Sciences, which has more recently merged to form the School of Psychology and Sport Science (approval number: P20-18/19).

Experimental approach

A two-year prospective observational design was implemented to identify noncontact lower limb injury risk factors in male senior academy Rugby Union players across two consecutive competitive seasons. Training load was quantified on each training and match-day, and further converted into Arbitrary Units (AU) as: mean weekly training load; mean change in weekly training load; difference between highest and lowest weekly training load; and maximum change in weekly training load [35, 36]. Furthermore, performance testing data were recorded at the start of each season for: aerobic fitness [37]; sprint times [38]; 1 repetition maximum (1RM) absolute and relative strength in the back squat [39]; and lower limb power via a counter movement jump (CMJ) [40] which were implemented and collected by the Strength and Conditioning Coach during the preseason. Each athlete was further screened at the

beginning of the season for previous injury history including previous concussions by the Team Physiotherapist. Musculoskeletal screening was administered at the beginning of each training day including measures of: ankle dorsiflexion angle; hamstring strength; adductor strength; bodyweight; and perceived muscle soreness ratings [41–45]. These measures were conducted and recorded by the Primary Researcher and later extracted to the team's athlete monitoring database. Non-contact lower limb injuries were noted by binary classification (injury yes/no), as well as further categorization into their specific location and severity (hip/groin, knee, hamstring, ankle, more than 28 days lost due to injury). All raw data points measured were then collated and standardised into one Excel database for pattern recognition analysis, to further examine the best performing predictive variables from agreeing algorithms and identifying athletes with greater likelihood of sustaining lower limb non-contact injuries based upon the monitoring variables.

Training load. Training and match loads were quantified using the session Rating of Perceived Exertion (SRPE) method using a 10-point category ratio scale CR-10) and the session or match duration [46]. Thirty minutes post session; the player self-reported their perceived exertion using the Borg-CR10 scale for each individual session (for example; Unit Skills, Gym, Rugby Training, Conditioning) which was subsequently multiplied by the duration of the session (minutes), expressed in arbitrary units (AU) for each player [47, 48]. Weekly means of training load were collated for each player throughout each season, including absolute week-to-week changes in training load, maximum change in week-to-week load over the season, and pre-season cumulative training load (sum each athlete's training load in the pre-season phase [49]).

Performance testing. Strength, speed, and aerobic fitness tests were performed during the pre-season phase to establish baseline performance. For this study, back squat 1RM was considered representative of lower limb strength and has previously identified as a predictive attribute for lower limb injuries [39]; athletes would perform a dynamic warm-up prior to testing, followed by utilizing an empty barbell to squat before gradually increasing the weight to 90% of their previous 1RM. Athletes were permitted three attempts from this threshold to establish their 1RM. Players were visually assessed by a UKSCA-certified strength and conditioning coach during each 1RM attempt to ensure appropriate technique was performed. Back squat 1RM absolute values were noted and then calculated relative to body mass for each player (1RM / body mass (kg)).

Player sprint speed was assessed over 10m and 40m from a stationery position. Athletes were permitted three attempts at maximum effort, with the fastest time being noted for analysis. Sprint speed was measured using a timing gate system (Brower Timing Systems, Utah, USA) using a single beam infrared light system with the data then sent directly to a handheld monitor. Inter-rater reliability of this procedure is considered at an inter-class correlation coefficient between 0.91–0.99 over 10m to 30m sprint tests [38].

The Bronco test [37] was administered to assess aerobic capacity and involved a shuttle-running time-trial over a total distance of 1200m. Distances of 0, 20, 40, and 60m were marked with cones and athletes were required to run from the 0-m to the 20-m marker and back, then run from the 0-m to the 40-m marker and back, and finally run from the 0-m to the 60-m marker and back. Completion of this sequence of shuttles was considered one repetition, with athletes completing five repetitions to cover 1200m as quickly as possible. Video-analysis was used to record the duration of each player's Bronco time. This test is commonly used in Rugby Union settings to measure maximal aerobic capacity and is considered a cost and time-efficient method for coaches to monitor high-intensity aerobic fitness [37].

A CMJ was used to establish lower limb explosive power, using a contact mat (Just Jump System, Probotics Inc., Huntsville, AL, USA). To measure jump height of a maximal effort jump from a static position; athletes would start on the contact mat in an upright position with

hands on hips followed by a dip into self-selected depth and would perform a vertical acceleration aiming to generate a maximal vertical jump height. During testing, athletes were permitted three attempts with 2-minute rest in-between each attempt, and the highest CMJ height was recorded for analysis.

Injury history. Previous injury history was noted as part of pre-season screening of the athletes and collected by the Team Physiotherapist. Injuries were further specified for each different injury location (groin, hamstring, knee, ankle) and previous injury history referred to a previous injury corresponding to the classified injury in the analysis. Additionally, if the athlete had sustained a concussion within the previous season, this was noted separately. Previous incidence of concussion has been found to increase the likelihood of subsequent soft-tissue injury in team sports [50, 51].

Musculoskeletal screening. Height (cm) and body mass (kg) were measured using a stadiometer (Bodycare Ltd, Warwickshire, UK) and digital platform scales (Model 705, Seca, Hamburg, Germany). Subjective measurements of muscle soreness for the lower limb were collected on a Likert scale of 1–10 prior to each training day [41]. A recording of 1 represented no soreness and 10 equalled very sore, and subjective markers have previously been deemed reliable and valid methods of recording muscle soreness [52]. Athletes would perform objective validated lower limb musculoskeletal screening tests prior to the commencement of each training day which would take place three times per week for the duration of the season, including: adductor squeeze (adductor strength; [42, 43]) dorsiflexion lunge angle with dominant to non-dominant leg comparison (ankle dorsiflexion range of movement; [44]) and hamstring squeeze with dominant to non-dominant leg comparison (hamstring strength; [45]). Detailed methods used for musculoskeletal screening are listed in [S1 Table](#).

Injury classifiers. Players were assigned class membership on six binary classification variables: non-contact lower limb injury of mild to moderate severity (lower limb injured vs non-injured class), non-contact hip/groin injury (hip/groin injured vs non-injured class), non-contact hamstring injury (hamstring injury vs non-injured class), non-contact knee injury (knee injury vs non-injured class), non-contact ankle injury (ankle injury vs non-injured class), and severe non-contact injury (severe injury vs non-severe/non-injury class). Injury data was collected by a designated Physiotherapist for this study and was defined as non-contact lower limb injury resulting in time-loss from training and/or match-play within the data collection period [53]. If an injury resulted in time-loss within this time-frame, the following training load and monitoring markers were omitted from the study until the player returned to full-training.

Statistical analysis

Bayesian pattern recognition analysis was performed collaboratively by RO and SE in R programming language (R Core Development, 2021), with the Tidyverse package used for data manipulation [54] and rWeka package used to compute WEKA machine learning algorithms [55, 56]. Bayesian pattern recognition analysis was performed a total of six times, once for each binary classification variable (non-contact lower limb injury, non-contact ankle injury, non-contact knee injury, non-contact hamstring injury, non-contact hip/groin injury and severe non-contact injury), to determine which of the 103 collected features (i.e., predictor variables such as eccentric hamstring strength discrepancy between limbs; see [Table 1](#) for a full list) best predicted group membership (i.e., injured or non-injured). In line with previous work utilising pattern recognition in sport [56, 57], each instance of the Bayesian pattern recognition analyses comprised of three stages: pre-processing, feature selection, and assessment of classification accuracy.

Pre-processing was performed at the beginning of each pattern recognition analysis to standardize features and operationalise classification variables. Specifically, raw data of features

Table 1. Potential predictive risk factor variable list.

Training Load				
Mean Weekly Training Load	Mean Change in Weekly Training Load	Highest to Lowest Weekly Training Load	Maximum Increase in Weekly Training Load	Preseason Cumulative Load
Performance Testing				
10m Sprint	40m Sprint	Bronco Fitness Test	Absolute Back Squat 1RM	Back Squat 1RM Relative to Bodyweight
Counter Movement Jump Test				
Injury History				
Previous Injury	Concussion in the Previous Season			
Musculoskeletal Screening				
Bodyweight	Stature	Age	Playing Position	
Mean Weekly Muscle Soreness	Mean Change in Weekly Muscle Soreness	Highest to Lowest Weekly Muscle Soreness	Maximum Increase in Weekly Muscle Soreness	
Mean Weekly Adductor Squeeze Test	Mean Change in Weekly Adductor Squeeze Test	Highest to Lowest Weekly Adductor Squeeze Test	Maximum Decrease in Weekly Adductor Squeeze Test	
Mean Weekly Ankle Dorsiflexion Angle	Mean Change in Weekly Ankle Dorsiflexion Angle	Highest to Lowest Weekly Ankle Dorsiflexion Angle	Maximum Decrease in Weekly Ankle Dorsiflexion Angle	Left to Right Leg Difference in Weekly Ankle Dorsiflexion Angle
Mean Weekly Hamstring Strength	Mean Change in Weekly Hamstring Strength	Highest to Lowest Weekly Hamstring Strength	Maximum Decrease in Weekly Hamstring Strength	Left to Right Leg Difference in Weekly Hamstring Strength

All 22 weekly variables were sub-divided into phases of the season [preseason, early, mid], resulting in 103 weekly variables entered into the feature selection stage.

<https://doi.org/10.1371/journal.pone.0307287.t001>

was z-scored using their mean and standard deviation, prior to being converted into a vector from 0 to 100 (with a player's score of 50 representing a score equivalent to the mean and a score of 60 representing a score 1SD above the mean, etc). Data of classification variables (i.e., injury incidence) was coded in a way where 0 represented an absence of injury for a player during the season, whilst 1 represented the presence of an injury.

Feature selection was performed following pre-processing, for each instance of pattern recognition. This involved the use of correlation attribute evaluator [58], gain ratio attribute evaluator [59], relief F attribute evaluator [60], and info gain attribute evaluator [59] to identify the features which best predicted group membership within the classification variable (i.e., 0 or 1). Features which were identified as being among the top-20 strongest predictors of group membership by at least two feature selection algorithms formed part of a resultant feature 'model' which proceeded to the next stage of analysis. Note, the present top-20 and minimum of 2 agreeing algorithm criterions were set in-line with previous pattern recognition research in sport [56, 57].

The classification accuracy (i.e., accuracy with which injured/non-injured group membership is determined) of the models that emerged from the feature selection stage was assessed via naïve bayes [61], J48 decision tree [62], support vector machine [63], and K-nearest neighbour [64] classification algorithms. These analyses were iterated using a 'leave one out' cross-validation procedure [65]. The resultant output from these analyses provided a mean of the four classification algorithms ratings of classification accuracy (in percentage terms), sensitivity (a measure of correctly identified members of the non-injured group), specificity (a measure of correctly identified members of the injured group) and area under the receiver operating characteristic (ROC) (an overall measure of model performance when used for classification). The process of pattern recognition specifically for Rugby Union is schematically represented in Fig 1.

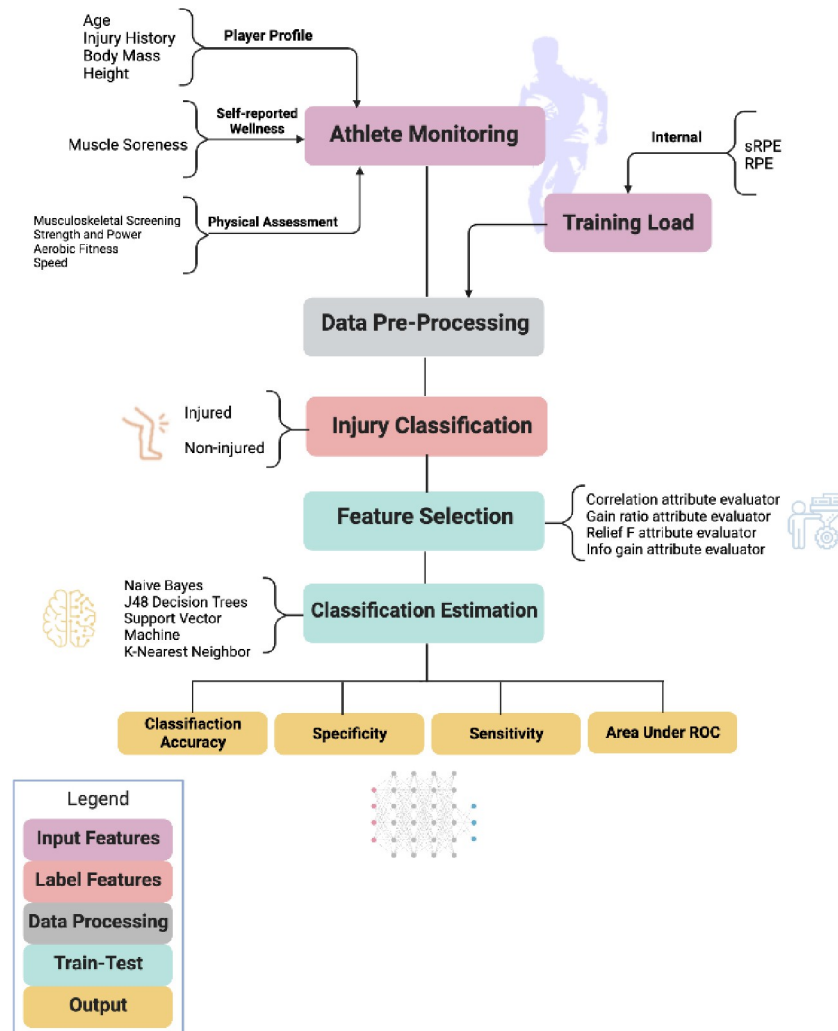


Fig 1. An adapted Figure [10] demonstrating the process of injury prediction validation using a pattern recognition approach. Athlete monitoring and training load refer to input variables for the pattern recognition algorithms. Injury classification is typically the binary response (injury yes/no) in labelling the training vectors. The models are those tested by the injury prediction algorithms in this study. Cross-validation and feature selection are processes of training, validating, and testing the models developed, further supported by the output determining the usefulness of the model. Furthermore, blue leaves describe how to train, validate, and test the model developed by the injury prediction algorithm.

<https://doi.org/10.1371/journal.pone.0307287.g001>

Results

Thirty-six athletes (Mean: age 20.7 ± 2.2 years, stature 183.3 ± 6.1 cm, body mass 98.9 ± 11.5 kg) participated in this study across two consecutive seasons (2018–19, 2019–20). Overall, there were 25 athletes who sustained lower limb non-contact injury (Mean: age 20.7 ± 2.2 years, stature 183.4 ± 6.6 cm, body mass 100.1 ± 12.3 kg), and 11 athletes not sustaining an injury (Mean: age 20.8 ± 2.4 years, stature 183.0 ± 5.3 cm, body mass 96.1 ± 9.4). The most common mechanism of lower limb non-contact injury was running (16% of all non-contact lower limb injuries), and the most injured location was the ankle (52% of non-contact lower limb injuries), with the most common diagnosis being ankle sprains (48%). [Table 2](#) contains the overall feature selection result, where models comprising of predictive risk factors were created for

Table 2. Algorithm agreements among feature selection models and their associated classification for non-contact lower limb injuries during the season.

Models												
Number of feature selection algorithms in agreement	Non-contact Lower Limb Injury				Non-contact Ankle Injury				Severe Non-Contact Lower Limb Injury			
	Training Load	Performance Testing	Injury History	Musculoskeletal Screening	Training Load	Performance Testing	Injury History	Musculoskeletal Screening	Training Load	Performance Testing	Injury History	Musculoskeletal Screening
4												Early season mean change in muscle soreness
3				Mid-season maximum reduction in adductor strength			Previous Ankle Injury					Position
2			Previous Lower Limb Injury	Mid-season mean change in dorsiflexion angle		10m Sprint 40m Sprint	Previous Concussion	Bodyweight Preseason dorsiflexion angle Early season maximum change in dorsiflexion angle	Highest to lowest difference in training load	10m Sprint 40m Sprint		Early season mean change in adductor strength Early season mean change in dorsiflexion angle Early season mean change in hamstring strength
Area Under the ROC	0.70				0.72				0.76			

<https://doi.org/10.1371/journal.pone.0307287.t002>

Table 3. Classification accuracy for severe non-contact lower limb injury during phases of the season.

Classifier	Classification Accuracy (%)	Sensitivity	Specificity	Area Under ROC Curve
Naïve Bayes	88.88	0.93	0.78	0.93
Support Vector Machine	77.77	1.00	0.12	0.55
K Nearest Neighbour	75.00	1.00	0.00	0.78
J48 Decision Tree	86.11	0.93	0.67	0.77
Mean	81.92	0.96	0.39	0.76

Sensitivity = 1 – true positive for injured group. Specificity = 1 – true positive for non-injured group. ROC = Receiver operating characteristic.

<https://doi.org/10.1371/journal.pone.0307287.t003>

Table 4. Classification accuracy for non-contact lower limb injury during phases of the season.

Classifier	Classification Accuracy [%]	Sensitivity	Specificity	Area Under ROC Curve
Naïve Bayes	75.00	0.46	0.88	0.76
Support Vector Machine	63.88	0.00	0.92	0.46
K Nearest Neighbour	72.22	0.19	0.96	0.77
J48 Decision Tree	83.33	0.55	0.96	0.79
Mean	73.6	0.30	0.93	0.70

Sensitivity = 1 – true positive for injured group. Specificity = 1 – true positive for non-injured group. ROC = Receiver operating characteristic.

<https://doi.org/10.1371/journal.pone.0307287.t004>

different injury types. Injury models producing areas under the ROC of 0.7 and above are discussed, as areas under the ROC of between 0.7 to 0.8 represent fair to moderate diagnostic accuracy [66]. Models for non-contact hip/groin, non-contact hamstring, and non-contact knee injuries produced areas under the ROC ranging between 0.39 to 0.61, therefore were excluded from the results for discussion. Given the inherent complexity of sport injury, achieving greater values of diagnostic accuracy is difficult within injury prediction [8], however

Table 5. Classification accuracy for non-contact ankle injury during phases of the season.

Classifier	Classification Accuracy (%)	Sensitivity	Specificity	Area Under ROC Curve
Naïve Bayes	77.77	0.84	0.64	0.75
Support Vector Machine	72.22	0.88	0.37	0.62
K Nearest Neighbour	80.55	0.96	0.46	0.78
J48 Decision Tree	72.2	0.76	0.64	0.72
Mean	75.7	0.86	0.52	0.71

Sensitivity = 1 – true positive for injured group. Specificity = 1 – true positive for non-injured group. ROC = Receiver operating characteristic.

<https://doi.org/10.1371/journal.pone.0307287.t005>

attention should be given to whether the predictive variables identified within the models are consistent with previous risk factor research and the applied implications of these findings. Pattern recognition analysis in this study identified models for severe non-contact lower limb injury, non-contact lower limb injury, and non-contact ankle injury with area under the ROC ranging between 0.70–0.76.

Table 3 shows a severe non-contact lower limb injury model from the classification analysis (Area under the ROC = 0.76). Players who sustained a severe non-contact injury saw less of a difference between their highest and lowest training load over the mid-season (2022.44 ± 397.36 AU vs 2406.59 ± 1266.86 AU), were slower in the 10m sprint (1.76 ± 0.14 s vs 1.69 ± 0.08 s), slower in the 40m sprint (5.44 ± 0.34 s vs 5.25 ± 0.29 s), saw greater muscle soreness increases in the early season (0.76 ± 1.24 vs 0.16 ± 0.56), greater reduction in adductor strength across the early season (-8.4 ± 15.65 mmHg vs -1.09 ± 3.27 mmhg), greater reduction in early season dorsiflexion angle ($-1.13 \pm 5.54^\circ$ vs $-0.07 \pm 0.56^\circ$), greater reduction in early season hamstring strength (-4.83 ± 17.54 mmHg vs -1.83 ± 4.99 mmHg) and were more likely to be forwards when compared to non-injured players.

Table 4 shows a non-contact lower limb injury model from the classification analysis (Area under the ROC = 0.70). Players who sustained a non-contact lower limb injury had a less reduction in adductor strength in the mid-season compared to non-injured players (-13.91 ± 10.78 mmHg vs -39.14 ± 31.05 mmhg), had a lower reduction in dorsiflexion angle in the mid-season compared to non-injured players ($-0.23 \pm 1.2^\circ$ vs $-0.70 \pm 1.7^\circ$) and had previously sustained a lower limb injury when compared to non-injured players.

Table 5 shows a non-contact ankle injury model from the classification analysis (Area under the ROC = 0.71). Players who sustained a non-contact ankle injury were slower in the 10m sprint (1.80 ± 0.15 s vs 1.69 ± 0.07 s), and the 40m sprint (5.43 ± 0.42 s vs 5.24 ± 0.23 s), had a lower dorsiflexion angle during the preseason ($32.45 \pm 11.06^\circ$ vs $37.08 \pm 15.82^\circ$), greater reduction in dorsiflexion angle over the early season phase ($-5.05 \pm 2.98^\circ$ vs $-3.24 \pm 1.93^\circ$), and were heavier in bodyweight (105.7 ± 15.2 kg vs 95.9 ± 8.1 kg) as well as having sustained a concussion within the previous season and a previous ankle injury when compared to non-injured players.

Discussion

The aim of this study was to identify the most predictive variables for non-contact lower limb injury in male Rugby Union players via a pattern recognition approach. We developed and evaluated predictive models comprised of features representing training load, performance testing, previous injury history and musculoskeletal screening markers collected throughout the phases of the season. Our findings not only successfully identify predictive models capturing the multifactorial nature of lower limb noncontact injury, but also explicitly specified thresholds for each of the features evaluated whereby there was an increased likelihood of injury; an invaluable tool for practitioners to utilise in implementing timely injury prevention strategies.

Severe non-contact lower limb injuries were attributed to smaller differences between highest to lowest midseason training loads, slower 10m and 40m sprint times, greater increase in muscle soreness in the early season, greater decrease in adductor strength in the early season, greater early season mean decrease in dorsiflexion angle, greater decrease in hamstring strength in the early season with forwards more likely to sustain an injury compared to backs. There is some evidence that a relationship exists between training load and injury risk in athletes [67–69], with underloading being a potential characteristic of predisposing athletes to non-contact injuries [70]. A smaller variation in weekly training loads within the mid-season

phase may suggest increased training monotony compared to the un-injured group, which may increase the risk of lower limb non-contact injury [71]. Additionally, enhanced sprinting ability has been considered a moderator for injury risk, and slower 10m and 40m sprint times increased injury risk in sub elite Rugby League players [72], reflective of our findings. Furthermore, isometric hamstring strength was measured in this study; reductions in hamstring strength was a predictor of severe lower limb non-contact injury, and previous studies do report that hamstring weakness predisposes athletes to injury [73, 74]. The severity of lower limb non-contact muscular strain injuries could be attributed to sudden activation of eccentric contractions of the hamstring muscles during the late swing phase of high-speed running [75–77]. Interventions for enhancing hamstring strength may therefore reduce lower limb injury severity in football and Rugby Union [78]. Tentatively from our findings, we may assume that greater training monotony, running exposure paired with limited sprinting ability and poor hamstring strength may increase the risk of severe lower limb non-contact injuries.

Self-reported measures of muscle soreness have been reported to be sensitive to changes in training load amongst team sports athletes [52], with our study indicating that greater increases in weekly perceived muscle soreness was symptomatic pre-lower limb injury. In recent machine learning approaches to injury risk within elite male volleyball, muscle soreness and training exertion predicted injury [79]. Additionally, reduced adductor strength and reduced dorsiflexion angle are associated with greater lower limb injury risk [18, 19, 80, 81], supporting these features as predictive variables for severe noncontact lower limb injury in this model. Restricted ankle dorsiflexion may also contribute to reduced hip and knee mobility, potentially increasing lower limb injury risk due to greater ground reaction forces through increased joint stiffness [18, 82]. Finally, we reported positional differences within the model with forwards at a greater risk of severe lower limb non-contact injury. Previous literature has stated trivial differences between injury severity for forwards versus backs [5, 83–85]. As the most common mechanism of non-contact lower limb injuries typically derive from running-related activities in Rugby Union [5, 85, 86], we can postulate that the greater risk for the forwards may be due to differing training stimulus compared with backs; backs are more likely to be involved in high-speed running activities in match-play and their training stimulus will replicate this [87–89], whereas the predominantly contact nature of the forwards position may lead to their game based training in preparation for high-speed running to be less developed [90]. In combination, the above features contribute to forwards with restricted strength and range of movement in their lower limb musculoskeletal screening, increased perception of muscle soreness, slower sprint times and greater mid-season training load monotony to have greater likelihood of sustaining severe non-contact lower limb injuries.

Non-contact lower limb injury risk factors included a lower change in maximum adductor strength in the mid-season, lower change in dorsiflexion angle in the mid-season mean and previous lower limb injury. Contrasting with previous findings in elite football that athletes with greater reductions in hip adductor strength were at an increased risk of lower limb injury [19, 80, 81]. As reduced absolute adductor strength has been found to increase injury risk in team sports, a possible explanation for our findings would be that the un-injured athletes in this analysis were 5% stronger than their injured counterparts at their preseason adductor strength baseline, potentially moderating their risk of lower limb non-contact injury. Similarly, it was identified that less reduction in ankle dorsiflexion was a risk factor for lower limb non-contact injury. As these results signify changes in the mid-season, it is important to consider natural decreases in flexibility over the course of typically congested match fixtures due to increased training loads and limited recovery [91]. Finally, previous lower limb injuries are considered primary risk factors for both recurrent injuries and new injuries in team sports [25–29], supporting the finding of this feature being a significant contributor to this injury

model. Despite some of these findings being atypical, it is important to note that models which are generated for specific diagnoses of lower extremity musculoskeletal injuries may be of greater sensitivity in multivariate modelling, producing more specific, interpretable findings for injury risk [92]. Future research within pattern recognition and injury risk should therefore be modelled on specific injury diagnosis to facilitate further specificity and generalisability of the models in an applied setting.

Non-contact ankle injuries were determined by lower mean dorsiflexion angle in the pre-season, reduced early season mean change in dorsiflexion angle, greater bodyweight, slower 10m and 40m sprint times, sustaining a concussion within the previous season, and history of previous ankle injury. Reductions in dorsiflexion angle is well established as having links to ankle and lower limb injury, by altering lower-extremity stiffness and forces generated through the ankle during landing mechanics [18]. Furthermore, heavier athletes in this study were at an increased risk of sustaining non-contact ankle injuries, with current evidence identifying greater body mass index as a predictor of ankle injury [8, 92]. Slower high-speed running ability was also identified as a predictor of ankle injury risk, and as previously discussed, reduced sprint speed has been associated with increased injury risk in team sports [16, 93]. Furthermore, the long-term effect of sustaining a concussion is being extensively researched in team sports, with some associating a previous concussion diagnosis to an increased risk of subsequent lower limb injury [50, 51], specifically lateral ankle sprains [94]. It is also important to consider that previous ankle sprains can result in significant reductions in ankle range of motion and chronic instability if not managed, with previous injury history being a strong risk factor for recurrence with regards to the clinical course of ankle sprains [95, 96]. Being a heavier athlete with a history of ankle and concussion injury combined with limited ankle dorsiflexion may not only increase joint loading, but also alter optimal movement patterns due to potential proprioceptive deficits following a concussion during activities such as running, landing, and cutting activities, creating this predisposition for re-injury in Rugby Union. By screening athletes in the preseason for ankle range of motion, sprinting ability and previous injury history, particularly previous concussion history, practitioners can implement timely and conservative interventions prior to season commencement to address discrepancies and reduce the risk of these preventable injuries.

Despite being able to identify injury prediction models for non-contact lower limb, non-contact ankle and severe non-contact injury, we were unable to substantiate injury models for knee, hamstring, or hip/groin injuries with adequate area under the ROC values of above 0.7. This is in line with previous literature stating that there were no identifiable risk factors for hamstring injuries in elite Australian Football using a similar statistical approach [97], however there is some evidence as to which variables could predict anterior cruciate ligament injuries in Basketball and Rugby [98, 99]. In future, incorporating more specific injury diagnosis could provide an opportunity for accurate injury prediction models to be formulated by pattern recognition algorithms, answering more complex questions beyond general injury risk [26].

Strength and limitations

The acknowledgement of pattern recognition's increasing acceptance in sport injury research [92, 97–99] reflects the realisation that injuries result from intricate interactions amongst modifiable and non-modifiable risk factors within a multidimensional context [12]. Recognising this complexity of injury causation, this study moves away from linear approaches which may oversimplify the composite nature of injuries, as approaches such as logistic regression do not handle imbalanced datasets that are typically seen within injury research [32]. Innovatively, this study adopted a dual approach, considering both absolute values of features and tracking

weekly changes, thus aligning with previous literature outlining the significance of monitoring changes in repeated measures for a more nuanced understanding of injury risk [100]. The breadth of data collected within this study offers a unique insight into the multifaceted nature of non-contact lower limb injuries within Rugby Union.

There were no specific psychosocial variables utilized as features for the models. Despite the use of subjective perception of muscle soreness in this study, previous studies have emphasised the importance of psychosocial well-being and injury risk [101], with documented models stating that psychological stressors, coping resources, life event stressors, personality and emotional state may be precursors for injury outcomes [102]. Despite the absence of explicit psychosocial variables, the present dataset distinguishes itself due to its thoroughness in comparison to other studies when exploring injury risk factors in sport. The data points collected seamlessly complemented usual monitoring processes and did not cause any additional time commitments for athletes. This positions the study as a robust exploration of non-contact injuries, capturing a comprehensive range of physiological factors contributing to injury risk whilst incorporating streamlined monitoring systems. Future research should include monitoring datapoints for psychosocial features in order to formulate a more holistic, multidisciplinary approach to sport injury.

As there are some differences and similarities in the current literature as to what predicts musculoskeletal non-contact injury in sport, it is important to consider differences between key performance indicators across sports which may alter an athlete's injury risk. An example of this may be the links between lower extremity stiffness and flexibility with lower limb injury risk, with a suggestion that there is an "optimal" level of flexibility depending on the key performance indicators of each sport [103]. To overcome this, baseline models from similar sports should be used to compare model outputs to create more generalisable results, as it has been recommended that comparing behaviours of injury prediction models built from different sports should be analysed both collectively and separately [32]. While this study represents and inaugural exploration of predictive variables for injury risk in Rugby Union, a comparative analysis with common features across our three models could serve as a step towards future generalisability. However, due to this being the first study to our knowledge examining predictive variables for non-contact injury risk in Rugby Union, this information was not available.

Conclusion

This is the first study to utilize pattern recognition analysis to examine the injury risk factors which pose the greatest risk to Rugby Union athletes in sustaining lower limb non-contact injuries. The present study provides a comprehensive approach to injury risk, accounting for their multifactorial nature and non-linear behaviours. Our findings indicate that training load parameters, performance testing results, athlete previous injury history and musculoskeletal screening combined can create holistic models for non-contact lower limb injury risk. Noteworthy predictive variables included differences in training load, reduced strength across musculoskeletal markers of adductor and hamstring strength, altered ankle dorsiflexion angles, perceived muscle soreness, slower sprint times, playing position and previous injury history. Appropriate weekly training load management, screening for strength and range of movement discrepancies across the season, enhancing sprinting abilities, and addressing previous injury history inclusive of previous concussions are some key considerations recommended for practitioners in formulating injury prevention strategies for non-contact lower limb injuries.

Supporting information

S1 Table. Daily musculoskeletal screening procedures.
(PDF)

Author Contributions

Conceptualization: Seren Lois Evans, Julian Owen.

Data curation: Seren Lois Evans, Gareth Whittaker, Oran Elphinstone Davis.

Formal analysis: Seren Lois Evans, Robin Owen.

Funding acquisition: Seren Lois Evans, Eleri Sian Jones, James Hardy, Julian Owen.

Investigation: Seren Lois Evans, Gareth Whittaker, Oran Elphinstone Davis.

Methodology: Seren Lois Evans, Julian Owen.

Project administration: Seren Lois Evans, Julian Owen.

Resources: Seren Lois Evans, Gareth Whittaker, Oran Elphinstone Davis, Julian Owen.

Software: Seren Lois Evans, Robin Owen.

Supervision: Seren Lois Evans, Eleri Sian Jones, James Hardy, Julian Owen.

Validation: Seren Lois Evans, Julian Owen.

Visualization: Seren Lois Evans, Robin Owen, Julian Owen.

Writing – original draft: Seren Lois Evans, Robin Owen, Julian Owen.

Writing – review & editing: Seren Lois Evans, Robin Owen, Gareth Whittaker, Oran Elphinstone Davis, Eleri Sian Jones, James Hardy, Julian Owen.

References

1. Brooks JHM. A prospective study of injuries and training amongst the England 2003 Rugby World Cup squad. *Br J Sports Med.* 2005 May 1; 39(5):288–93. <https://doi.org/10.1136/bjsm.2004.013391> PMID: 15849293
2. Chalmers DJ, Samaranyaka A, Gulliver P, McNoe B. Risk factors for injury in rugby union football in New Zealand: a cohort study. *Br J Sports Med.* 2012 Feb; 46(2):95–102. <https://doi.org/10.1136/bjsports-2011-090272> PMID: 22190761
3. Evans SL, Davis OE, Jones ES, Hardy J, Owen JA. Match and training injury risk in semi-professional rugby union: A four-year study. *J Sci Med Sport.* 2022 Jan;S1440244022000093.
4. Roberts SP, Trewartha G, England M, Shaddock G, Stokes KA. Epidemiology of time-loss injuries in English community-level rugby union. *BMJ Open.* 2013 Nov; 3(11):e003998. <https://doi.org/10.1136/bmjopen-2013-003998> PMID: 24240143
5. Fuller C, Taylor A, Douglas M, Raftery M. Rugby World Cup 2019 injury surveillance study. *South Afr J Sports Med.* 2020 May 4; 32(1):1–6. <https://doi.org/10.17159/2078-516X/2020/v32i1a8062> PMID: 36818969
6. Williams S, Trewartha G, Kemp SPT, Brooks JHM, Fuller CW, Taylor AE, et al. Time loss injuries compromise team success in Elite Rugby Union: a 7-year prospective study. *Br J Sports Med.* 2016 Jun; 50(11):651–6. <https://doi.org/10.1136/bjsports-2015-094798> PMID: 26552415
7. Drew MK, Raysmith BP, Charlton PC. Injuries impair the chance of successful performance by sportspeople: a systematic review. *Br J Sports Med.* 2017 Aug; 51(16):1209–14. <https://doi.org/10.1136/bjsports-2016-096731> PMID: 28446456
8. Jauhainen S, Kauppi JP, Leppänen M, Pasanen K, Parkkari J, Vasankari T, et al. New Machine Learning Approach for Detection of Injury Risk Factors in Young Team Sport Athletes. *Int J Sports Med.* 2021 Feb; 42(02):175–82. <https://doi.org/10.1055/a-1231-5304> PMID: 32920800

9. Bahr R. Why screening tests to predict injury do not work—and probably never will. . . : a critical review. *Br J Sports Med.* 2016 Jul; 50(13):776–80.
10. Rossi A, Pappalardo L, Cintia P. A Narrative Review for a Machine Learning Application in Sports: An Example Based on Injury Forecasting in Soccer. *Sports.* 2021 Dec 24; 10(1):5. <https://doi.org/10.3390/sports10010005> PMID: 35050970
11. Shmueli G. To Explain or to Predict? *Stat Sci [Internet].* 2010 Aug 1 [cited 2022 Oct 10]; 25(3). Available from: <https://projecteuclid.org/journals/statistical-science/volume-25/issue-3/To-Explain-or-to-Predict/10.1214/10-STS330.full>
12. Bittencourt NFN, Meeuwisse WH, Mendonça LD, Nettel-Aguirre A, Ocarino JM, Fonseca ST. Complex systems approach for sports injuries: moving from risk factor identification to injury pattern recognition—narrative review and new concept. *Br J Sports Med.* 2016 Nov; 50(21):1309–14. <https://doi.org/10.1136/bjsports-2015-095850> PMID: 27445362
13. Mooney M, Charlton PC, Soltanzadeh S, Drew MK. Who ‘owns’ the injury or illness? Who ‘owns’ performance? Applying systems thinking to integrate health and performance in elite sport. *Br J Sports Med.* 2017 Jul; 51(14):1054–5. <https://doi.org/10.1136/bjsports-2016-096649> PMID: 28330847
14. Tee JC, Bekker S, Collins R, Klingbiel J, van Rooyen I, van Wyk D, et al. The efficacy of an iterative “sequence of prevention” approach to injury prevention by a multidisciplinary team in professional rugby union. *J Sci Med Sport.* 2018 Sep; 21(9):899–904. <https://doi.org/10.1016/j.jsams.2018.02.003> PMID: 29544911
15. Duhig S, Shield AJ, Opar D, Gabbett TJ, Ferguson C, Williams M. Effect of high-speed running on hamstring strain injury risk. *Br J Sports Med.* 2016 Dec; 50(24):1536–40. <https://doi.org/10.1136/bjsports-2015-095679> PMID: 27288515
16. Malone S, Owen A, Mendes B, Hughes B, Collins K, Gabbett TJ. High-speed running and sprinting as an injury risk factor in soccer: Can well-developed physical qualities reduce the risk? *J Sci Med Sport.* 2018 Mar; 21(3):257–62. <https://doi.org/10.1016/j.jsams.2017.05.016> PMID: 28595870
17. Backman LJ, Danielson P. Low Range of Ankle Dorsiflexion Predisposes for Patellar Tendinopathy in Junior Elite Basketball Players: A 1-Year Prospective Study. *Am J Sports Med.* 2011 Dec; 39(12):2626–33. <https://doi.org/10.1177/0363546511420552> PMID: 21917610
18. Mason-Mackay AR, Whatman C, Reid D. The effect of reduced ankle dorsiflexion on lower extremity mechanics during landing: A systematic review. *J Sci Med Sport.* 2017 May; 20(5):451–8. <https://doi.org/10.1016/j.jsams.2015.06.006> PMID: 26117159
19. Moreno-Pérez V, Travassos B, Calado A, Gonzalo-Skok O, Del Coso J, Mendez-Villanueva A. Adductor squeeze test and groin injuries in elite football players: A prospective study. *Phys Ther Sport.* 2019 May; 37:54–9. <https://doi.org/10.1016/j.ptsp.2019.03.001> PMID: 30856592
20. Tyler TF, Nicholas SJ, Campbell RJ, McHugh MP. The Association of Hip Strength and Flexibility with the Incidence of Adductor Muscle Strains in Professional Ice Hockey Players. *Am J Sports Med.* 2001 Mar; 29(2):124–8. <https://doi.org/10.1177/03635465010290020301> PMID: 11292035
21. Tiernan C, Lyons M, Comyns T, Nevill AM, Warrington G. The Relationship Between Adductor Squeeze Strength, Subjective Markers of Recovery and Training Load in Elite Rugby Players. *J Strength Cond Res.* 2019 Nov; 33(11):2926–31. <https://doi.org/10.1519/JSC.0000000000003370> PMID: 31469760
22. Croisier JL, Ganteaume S, Binet J, Genty M, Ferret JM. Strength Imbalances and Prevention of Hamstring Injury in Professional Soccer Players: A Prospective Study. *Am J Sports Med.* 2008 Aug; 36(8):1469–75. <https://doi.org/10.1177/0363546508316764> PMID: 18448578
23. Henderson G, Barnes CA, Portas MD. Factors associated with increased propensity for hamstring injury in English Premier League soccer players. *J Sci Med Sport.* 2010 Jul; 13(4):397–402. <https://doi.org/10.1016/j.jsams.2009.08.003> PMID: 19800844
24. van Dyk N, Bahr R, Burnett AF, Whiteley R, Bakken A, Mosler A, et al. A comprehensive strength testing protocol offers no clinical value in predicting risk of hamstring injury: a prospective cohort study of 413 professional football players. *Br J Sports Med.* 2017 Dec; 51(23):1695–702. <https://doi.org/10.1136/bjsports-2017-097754> PMID: 28756392
25. Hagglund M. Previous injury as a risk factor for injury in elite football: a prospective study over two consecutive seasons. *Br J Sports Med.* 2006 Jul 6; 40(9):767–72. <https://doi.org/10.1136/bjism.2006.026609> PMID: 16855067
26. Luu BC, Wright AL, Haerberle HS, Karnuta JM, Schickendantz MS, Makhni EC, et al. Machine learning outperforms logistic regression analysis to predict next-season NHL player injury: an analysis of 2322 players from 2007 to 2017. *Orthopaedic journal of sports medicine.* 2020 Sep 25; 8(9):2325967120953404. <https://doi.org/10.1177/2325967120953404> PMID: 33029545

27. Pasanen K, Rossi MT, Parkkari J, Heinonen A, Steffen K, Myklebust G, et al. Predictors of lower extremity injuries in team sports (PROFITS-study): a study protocol. *BMJ Open Sport Exerc Med*. 2015 Dec; 1(1):e000076. <https://doi.org/10.1136/bmjsem-2015-000076> PMID: 27900143
28. Toohey LA, Drew MK, Cook JL, Finch CF, Gaida JE. Is subsequent lower limb injury associated with previous injury? A systematic review and meta-analysis. *Br J Sports Med*. 2017 Dec; 51(23):1670–8. <https://doi.org/10.1136/bjsports-2017-097500> PMID: 28784622
29. Walden M. High risk of new knee injury in elite footballers with previous anterior cruciate ligament injury * Commentary. *Br J Sports Med*. 2006 Feb 1; 40(2):158–62.
30. Sigurdson H, Chan J. Machine Learning Applications to Sports Injury: A Review: In: Proceedings of the 9th International Conference on Sport Sciences Research and Technology Support [Internet]. Online Streaming,—Select a Country—: SCITEPRESS—Science and Technology Publications; 2021 [cited 2022 Apr 28]. p. 157–68. <https://www.scitepress.org/DigitalLibrary/Link.aspx?doi=10.5220/0010717100003059>
31. Galar M, Fernández A, Barrenechea E, Herrera F. EUSBoost: Enhancing ensembles for highly imbalanced data-sets by evolutionary undersampling. *Pattern Recognit*. 2013 Dec; 46(12):3460–71.
32. López-Valenciano A, Ayala F, Puerta Jos M, De Ste Croix MBA, Vera-Garcia FJ, Hernández-Sánchez S, et al. A Preventive Model for Muscle Injuries: A Novel Approach based on Learning Algorithms. *Med Sci Sports Exerc*. 2018 May; 50(5):915–27. <https://doi.org/10.1249/MSS.0000000000001535> PMID: 29283933
33. Rommers N, Rössler R, Verhagen E, Vandecasteele F, Verstockt S, Vaeyens R, et al. A Machine Learning Approach to Assess Injury Risk in Elite Youth Football Players. *Med Sci Sports Exerc*. 2020 Aug; 52(8):1745–51. <https://doi.org/10.1249/MSS.0000000000002305> PMID: 32079917
34. The Welsh Rugby Union National League Rules 2022–2023.
35. Dubois R, Paillard T, McGrath D, Chamari K, Maurelli O, Polly S, et al. Changes in training load, running performance, lower body power and biochemical characteristics of back players throughout a professional Rugby Union season. *J Hum Sport Exerc* [Internet]. 2017 [cited 2020 Jul 9]; 12(1). <http://hdl.handle.net/10045/66788>
36. Cross MJ, Williams S, Trewartha G, Kemp SPT, Stokes KA. The Influence of In-Season Training Loads on Injury Risk in Professional Rugby Union. *Int J Sports Physiol Perform*. 2016 Apr; 11(3):350–5. <https://doi.org/10.1123/ijsp.2015-0187> PMID: 26309331
37. Berthon P, Fellmann N, Bedu M, Beaune B, Dabonneville M, Coudert J, et al. A 5-min running field test as a measurement of maximal aerobic velocity. *Eur J Appl Physiol*. 1997 Feb 1; 75(3):233–8.
38. Shalfawi SAI, Enoksen E, Tønnessen E, Ingebrigtsen J. Assessing test-retest reliability of the portable brower speed trap ii testing system. 2012; 7.
39. Case MJ, Knudson DV, Downey DL. Barbell Squat Relative Strength as an Identifier for Lower Extremity Injury in Collegiate Athletes. *J Strength Cond Res*. 2020 May; 34(5):1249–53. <https://doi.org/10.1519/JSC.0000000000003554> PMID: 32084107
40. Furlong LAM, Harrison AJ, Jensen RL. Measures of Strength and Jump Performance Can Predict 30-m Sprint Time in Rugby Union Players. *J Strength Cond Res* [Internet]. 2019 Apr 17 [cited 2021 May 28]; Publish Ahead of Print. Available from: <https://journals.lww.com/00124278-900000000-94840>
41. Montgomery PG, Hopkins WG. The effects of game and training loads on perceptual response of muscle soreness in Australian football. *Int J Sports Physiol Perform*. 2013; 8(3):312–8.
42. Toohey LA, De Noronha M, Taylor C, Thomas J. Is a sphygmomanometer a valid and reliable tool to measure the isometric strength of hip muscles? A systematic review. *Physiother Theory Pract*. 2015 Feb 17; 31(2):114–9. <https://doi.org/10.3109/09593985.2014.963905> PMID: 25286193
43. Toohey LA, de Noronha M, Taylor C, Thomas J. The validity and reliability of the sphygmomanometer for hip strength assessment in Australian football players. *Physiother Theory Pract*. 2018 Feb 1; 34(2):131–6. <https://doi.org/10.1080/09593985.2017.1374492> PMID: 28891759
44. Banwell HA, Uden H, Marshall N, Altmann C, Williams CM. The iPhone Measure app level function as a measuring device for the weight bearing lunge test in adults: a reliability study. *J Foot Ankle Res*. 2019 Dec; 12(1):37. <https://doi.org/10.1186/s13047-019-0347-9> PMID: 31333763
45. Mondin D, Owen JA, Negro M, D'Antona G. Validity and Reliability of a Non-invasive Test to Assess Quadriceps and Hamstrings Strength in Athletes. *Front Physiol*. 2018 Nov 29; 9:1702. <https://doi.org/10.3389/fphys.2018.01702> PMID: 30555343
46. Borg G. Borg's perceived exertion and pain scales. *Human kinetics*; 1998.
47. Haddad M, Stylianides G, Djaoui L, Dellal A, Chamari K. Session-RPE Method for Training Load Monitoring: Validity, Ecological Usefulness, and Influencing Factors. *Front Neurosci*. 2017 Nov 2; 11:612. <https://doi.org/10.3389/fnins.2017.00612> PMID: 29163016

48. Scantlebury S, Till K, Sawczuk T, Phibbs P, Jones B. Validity of Retrospective Session Rating of Perceived Exertion to Quantify Training Load in youth Athletes: *J Strength Cond Res*. 2018 Jul; 32(7):1975–80. <https://doi.org/10.1519/JSC.0000000000002099> PMID: 29939948
49. Evans SL, Whittaker G, Elphinstone Davis O, Jones ES, Hardy J, Owen JA. Noncontact Injury Distribution and Relationship With Preseason Training Load and Nonmodifiable Risk Factors in Rugby Union Players Across Multiple Seasons. *J Strength Cond Res*. 2023 Jul; 37(7):1456–62. <https://doi.org/10.1519/JSC.0000000000004402> PMID: 36445853
50. Hunzinger KJ, Costantini KM, Swanik CB, Buckley TA. Diagnosed concussion is associated with increased risk for lower extremity injury in community rugby players. *J Sci Med Sport*. 2021 Apr; 24(4):368–72. <https://doi.org/10.1016/j.jsams.2020.10.013> PMID: 33176983
51. Nordström A, Nordström P, Ekstrand J. Sports-related concussion increases the risk of subsequent injury by about 50% in elite male football players. *Br J Sports Med*. 2014 Oct; 48(19):1447–50. <https://doi.org/10.1136/bjsports-2013-093406> PMID: 25082616
52. Gallo TF, Cormack SJ, Gabbett TJ, Lorenzen CH. Self-Reported Wellness Profiles of Professional Australian Football Players During the Competition Phase of the Season. *J Strength Cond Res*. 2017 Feb; 31(2):495–502. <https://doi.org/10.1519/JSC.0000000000001515> PMID: 27243912
53. Fuller CW, Molloy MG, Bagate C, Bahr R, Brooks JHM, Donson H, et al. Consensus statement on injury definitions and data collection procedures for studies of injuries in rugby union. *Br J Sports Med*. 2007 Jan 29; 41(5):328–31. <https://doi.org/10.1136/bjism.2006.033282> PMID: 17452684
54. Wickham H, Wickham M. Package 'tidyverse' [Internet]. 2019. <http://tidyverse.tidyverse.org>.
55. Hornik K, Buchta C, Zeileis A. Open-source machine learning: R meets Weka. *Comput Stat*. 2009 May; 24(2):225–32.
56. Jones BD, Hardy L, Lawrence G, Kuncheva LI, Preez TD, Brandon R, et al. The Identification of “Game Changers” in England Cricket’s Developmental Pathway for Elite Spin Bowling: A Machine Learning Approach. 2019; 2(2):30.
57. Owen J, Owen R, Hughes J, Leach J, Anderson D, Jones E. Psychosocial and Physiological Factors Affecting Selection to Regional Age-Grade Rugby Union Squads: A Machine Learning Approach. *Sports*. 2022 Feb 28; 10(3):35. <https://doi.org/10.3390/sports10030035> PMID: 35324643
58. Hall MA. Correlation-based Feature Selection for Machine Learning. 1999; 199.
59. Gnanambal S, Thangaraj M, Meenatchi VT, Gayathri V. Classification algorithms with attribute selection: an evaluation study using WEKA. *Int J Adv Netowrking Appl*. 2018; 9(6):3640–4.
60. Kira K, Rendell L. The feature selection problem: Traditional methods and a new algorithm. *Aaai*. 1992; 2(1992a):129–34.
61. John GH, Langley P. Static Versus Dynamic Sampling for Data Mining. 1994; 4.
62. Quinlan JR. Programs for machine learning. San Mateo, California: Morgan Kaufmann Publishers; 1993.
63. Keerthi SS, Chapelle O, DeCoste D. Building Support Vector Machines with Reduced Classifier Complexity. 2006; 23.
64. Aha D, Kibler DW, Albert MK. Instance-based learning algorithms. *Mach Learn*. 1991; 6:37–66.
65. Vehtari A, Gelman A, Gabry J. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat Comput*. 2017 Sep; 27(5):1413–32.
66. Obuchowski NA, Lieber ML, Wians FH. ROC Curves in Clinical Chemistry: Uses, Misuses, and Possible Solutions. *Clin Chem*. 2004 Jul 1; 50(7):1118–25. <https://doi.org/10.1373/clinchem.2004.031823> PMID: 15142978
67. Drew MK, Finch CF. The Relationship Between Training Load and Injury, Illness and Soreness: A Systematic and Literature Review. *Sports Med*. 2016 Jun; 46(6):861–83. <https://doi.org/10.1007/s40279-015-0459-8> PMID: 26822969
68. Eckard TG, Padua DA, Hearn DW, Pexa BS, Frank BS. The Relationship Between Training Load and Injury in Athletes: A Systematic Review. *Sports Med*. 2018 Aug; 48(8):1929–61. <https://doi.org/10.1007/s40279-018-0951-z> PMID: 29943231
69. Jones CM, Griffiths PC, Mellalieu SD. Training Load and Fatigue Marker Associations with Injury and Illness: A Systematic Review of Longitudinal Studies. *Sports Med*. 2017 May; 47(5):943–74. <https://doi.org/10.1007/s40279-016-0619-5> PMID: 27677917
70. Gabbett TJ, Kennelly S, Sheehan J, Hawkins R, Milsom J, King E, et al. If overuse injury is a ‘training load error’, should undertraining be viewed the same way?. *British Journal of Sports Medicine*. 2016 Sep 1; 50(17):1017–8. <https://doi.org/10.1136/bjsports-2016-096308> PMID: 27251895
71. Delecroix B, Mccall A, Dawson B, Berthoin S, Dupont G. Workload monotony, strain and non-contact injury incidence in professional football players. *Sci Med Footb*. 2019 Apr 3; 3(2):105–8.

72. Gabbett TJ, Domrow N. Risk Factors for Injury in Subelite Rugby League Players. *Am J Sports Med.* 2005 Mar; 33(3):428–34. <https://doi.org/10.1177/0363546504268407> PMID: 15716260
73. Bourne MN, Opar DA, Williams MD, Shield AJ. Eccentric Knee Flexor Strength and Risk of Hamstring Injuries in Rugby Union: A Prospective Study. *Am J Sports Med.* 2015 Nov; 43(11):2663–70. <https://doi.org/10.1177/0363546515599633> PMID: 26337245
74. Fousekis K, Tsepis E, Vagenas G. Intrinsic Risk Factors of Noncontact Ankle Sprains in Soccer: A Prospective Study on 100 Professional Players. *Am J Sports Med.* 2012 Aug; 40(8):1842–50. <https://doi.org/10.1177/0363546512449602> PMID: 22700889
75. Danielsson A, Horvath A, Senorski C, Alentorn-Geli E, Garrett WE, Cugat R, et al. The mechanism of hamstring injuries—a systematic review. *BMC Musculoskelet Disord.* 2020 Dec; 21(1):641. <https://doi.org/10.1186/s12891-020-03658-8> PMID: 32993700
76. Liu H, Garrett WE, Moorman CT, Yu B. Injury rate, mechanism, and risk factors of hamstring strain injuries in sports: A review of the literature. *J Sport Health Sci.* 2012 Sep; 1(2):92–101.
77. Lovering RM, Hakim M, Moorman CT, De Deyne PG. The contribution of contractile pre-activation to loss of function after a single lengthening contraction. *J Biomech.* 2005 Jul; 38(7):1501–7. <https://doi.org/10.1016/j.jbiomech.2004.07.008> PMID: 15922761
78. Brooks JHM, Fuller CW, Kemp SPT, Reddin DB. Incidence, Risk, and Prevention of Hamstring Muscle Injuries in Professional Rugby Union. *Am J Sports Med.* 2006 Aug; 34(8):1297–306. <https://doi.org/10.1177/0363546505286022> PMID: 16493170
79. de Leeuw AW, van der Zwaard S, van Baar R, Knobbe A. Personalized machine learning approach to injury monitoring in elite volleyball players. *Eur J Sport Sci.* 2022 Apr 3; 22(4):511–20. <https://doi.org/10.1080/17461391.2021.1887369> PMID: 33568023
80. Markovic G, Šarabon N, Pausic J, Hadžić V. Adductor Muscles Strength and Strength Asymmetry as Risk Factors for Groin Injuries among Professional Soccer Players: A Prospective Study. *Int J Environ Res Public Health.* 2020 Jul 9; 17(14):4946. <https://doi.org/10.3390/ijerph17144946> PMID: 32659937
81. Whittaker JL, Small C, Maffey L, Emery CA. Risk factors for groin injury in sport: an updated systematic review. *Br J Sports Med.* 2015 Jun; 49(12):803–9. <https://doi.org/10.1136/bjsports-2014-094287> PMID: 25833903
82. Fong DTP, Hong Y, Chan LK, Yung PSH, Chan KM. A Systematic Review on Ankle Injury and Ankle Sprain in Sports: *Sports Med.* 2007; 37(1):73–94. <https://doi.org/10.2165/00007256-200737010-00006> PMID: 17190537
83. Bathgate A. A prospective study of injuries to elite Australian rugby union players * Commentary. *Br J Sports Med.* 2002 Aug 1; 36(4):265–9.
84. Bitchell CL, Mathema P, Moore IS. Four-year match injury surveillance in male Welsh professional Rugby Union teams. *Phys Ther Sport.* 2020 Mar; 42:26–32. <https://doi.org/10.1016/j.ptsp.2019.12.001> PMID: 31869752
85. Williams S, Trewartha G, Kemp S, Stokes K. A Meta-Analysis of Injuries in Senior Men's Professional Rugby Union. *Sports Med.* 2013 Oct; 43(10):1043–55. <https://doi.org/10.1007/s40279-013-0078-1> PMID: 23839770
86. West SW, Starling L, Kemp S, Williams S, Cross M, Taylor A, et al. Trends in match injury risk in professional male rugby union: a 16-season review of 10 851 match injuries in the English Premiership (2002–2019): the Professional Rugby Injury Surveillance Project. *Br J Sports Med.* 2021 Jun; 55(12):676–82. <https://doi.org/10.1136/bjsports-2020-102529> PMID: 33046453
87. Cahill N, Lamb K, Worsfold P, Headey R, Murray S. The movement characteristics of English Premiership rugby union players. *J Sports Sci.* 2013 Feb; 31(3):229–37. <https://doi.org/10.1080/02640414.2012.727456> PMID: 23009129
88. Cunniffe B, Proctor W, Baker JS, Davies B. An Evaluation of the Physiological Demands of Elite Rugby Union Using Global Positioning System Tracking Software: *J Strength Cond Res.* 2009 Jul; 23(4):1195–203. <https://doi.org/10.1519/JSC.0b013e3181a3928b> PMID: 19528840
89. Roberts SP, Trewartha G, Higgitt RJ, El-Abd J, Stokes KA. The physical demands of elite English rugby union. *J Sports Sci.* 2008 Jun; 26(8):825–33. <https://doi.org/10.1080/02640410801942122> PMID: 18569548
90. Tee JC, Lambert MI, Coopoo Y. GPS comparison of training activities and game demands of professional rugby union. *Int J Sports Sci Coach.* 2016 Apr; 11(2):200–11.
91. Moreno-Pérez V, Soler A, Ansa A, López-Samanes Á, Madruga-Parera M, Beato M, et al. Acute and chronic effects of competition on ankle dorsiflexion ROM in professional football players. *Eur J Sport Sci.* 2020 Jan 2; 20(1):51–60. <https://doi.org/10.1080/17461391.2019.1611930> PMID: 31072261

92. Henriquez M, Sumner J, Faherty M, Sell T, Bent B. Machine learning to predict lower extremity musculoskeletal injury risk in student athletes. *Frontiers in sports and active living*. 2020 Nov 19; 2:576655. <https://doi.org/10.3389/fspor.2020.576655> PMID: 33345141
93. de la Motte SJ, Lisman P, Gribbin TC, Murphy K, Deuster PA. Systematic Review of the Association Between Physical Fitness and Musculoskeletal Injury Risk: Part 3—Flexibility, Power, Speed, Balance, and Agility. *J Strength Cond Res*. 2019 Jun; 33(6):1723–35. <https://doi.org/10.1519/JSC.0000000000002382> PMID: 29239989
94. Gilbert FC, Burdette GT, Joyner AB, Llewellyn TA, Buckley TA. Association Between Concussion and Lower Extremity Injuries in Collegiate Athletes. *Sports Health Multidiscip Approach*. 2016 Nov; 8(6):561–7. <https://doi.org/10.1177/1941738116666509> PMID: 27587598
95. Mattacola CG, Dwyer MK. Rehabilitation of the Ankle After Acute Sprain or Chronic Instability.: 17.
96. van Rijn RM, van Os AG, Bernsen RMD, Luijsterburg PA, Koes BW, Bierma-Zeinstra SMA. What Is the Clinical Course of Acute Ankle Sprains? A Systematic Literature Review. *Am J Med*. 2008 Apr; 121(4):324–331.e7. <https://doi.org/10.1016/j.amjmed.2007.11.018> PMID: 18374692
97. Ruddy JD, Shield AJ, Maniar N, Williams MD, Duhig S, Timmins RG, et al. Predictive Modeling of Hamstring Strain Injuries in Elite Australian Footballers. *Med Sci Sports Exerc*. 2018 May; 50(5):906–14. <https://doi.org/10.1249/MSS.0000000000001527> PMID: 29266094
98. Taborri J, Molinaro L, Santospagnuolo A, Vetrano M, Vulpiani MC, Rossi S. A Machine-Learning Approach to Measure the Anterior Cruciate Ligament Injury Risk in Female Basketball Players. *Sensors*. 2021 Apr 30; 21(9):3141.
99. Tedesco S, Crowe C, Ryan A, Sica M, Scheurer S, Clifford AM, 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*. 2020 May 27; 20(11):3029. <https://doi.org/10.3390/s20113029> PMID: 32471051
100. Meeuwisse WH, Tyreman H, Hagel B, Emery C. A Dynamic Model of Etiology in Sport Injury: The Recursive Nature of Risk and Causation: *Clin J Sport Med*. 2007 May; 17(3):215–9. <https://doi.org/10.1097/JSM.0b013e3180592a48> PMID: 17513916
101. Johnson U. Athletes' experiences of psychosocial risk factors preceding injury. *Qual Res Sport Exerc Health*. 2011 Mar; 3(1):99–115.
102. Ivarsson A, Johnson U, Andersen MB, Tranaeus U, Stenling A, Lindwall M. Psychosocial Factors and Sport Injuries: Meta-analyses for Prediction and Prevention. *Sports Med*. 2017 Feb; 47(2):353–65. <https://doi.org/10.1007/s40279-016-0578-x> PMID: 27406221
103. Brazier J, Maloney S, Bishop C, Read PJ, Turner AN. Lower Extremity Stiffness: Considerations for Testing, Performance Enhancement, and Injury Risk. *J Strength Cond Res*. 2019 Apr; 33(4):1156–66. <https://doi.org/10.1519/JSC.0000000000002283> PMID: 29112054