



Article

Can We Predict Adductor Strain? A Predictive Analysis of a Major League Soccer (MLS) Cohort Spanning from 2019 to 2022

Rebecca Davis ^{1,2,*} , Benjamin C. Brewer ³, Martha Hall ¹ and Jill S. Higginson ²

¹ Biomechanics and Movement Science, University of Delaware, Newark, DE 19716, USA

² Mechanical Engineering Department, University of Delaware, Newark, DE 19716, USA

³ Biostatistics, Department of Epidemiology, University of Delaware, Newark, DE 19716, USA

* Correspondence: davisrl@udel.edu

Abstract

Background: Despite the high prevalence of adductor injury in soccer, there is limited injury-specific predictive modeling to identify common risk factors. The objective of this study was to create an adductor strain prediction model utilizing injury, game, and performance data collected from a cohort of professional Major League Soccer (MLS) players. **Methods:** We identified potential risk factors for soft tissue, non-contact adductor strain using a predictive machine learning model framework. Performance and injury data were collected between the 2019 to 2022 seasons of one professional MLS team. We utilized Random Forest (RF) machine learning models with Synthetic Minority Oversampling (SMOTE) to predict soft tissue, non-contact adductor strain injury amongst the cohort. Features chosen to be implemented in the model included injury, game, and performance data. **Results:** From the four models constructed in this study, the best performing model included Catapult Global Position System (GPS)/Internal Measurement Unit (IMU), strength, injury, and game data using a weekly structure determined by F1 score. Multiple models indicated that not having a previous injury lowers the odds of a future injury in the following week or month. Forwards had greater odds of injury whereas defenders had lower odds of injury. Greater hamstring max force lowered odds of injury whereas a greater amount of change of direction efforts increased the odds of injury in the following week or month. Adductor-to-abductor max strength ratio showed conflicting results regarding the odds of future injury. **Conclusions:** Through the utilization of RF and SMOTE, we were able to successfully predict adductor injuries in an MLS cohort utilizing injury, game, and performance metrics. Validation in a larger cohort would be highly recommended before utilizing the findings of this study in the design of injury prevention protocols.

Keywords: adductor strain; predictive modeling; professional soccer; MLS; injury risk; model framework



Academic Editor: Federico Roggio

Received: 7 January 2025

Revised: 12 February 2026

Accepted: 16 February 2026

Published: 5 March 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Adductor strains are one of the most common injuries endured by soccer players at all levels [1–4]. Depending on the grade of strain, rehabilitation, and/or the surgical procedures undergone, athletes can experience anywhere from a single day to an entire season out of play, however, the median time away from play ranges from 14 to 22 days [4,5]. Adductor reinjuries have been found to occur in 6 to 21% of injured adult male soccer players ranging in competitive levels [5–7] and have been found to have a significantly longer recovery time compared to other lower extremity injuries [5]. Due to the impact

this can have on a professional athlete's career and team, the scientific community has made it a priority to identify risk factors for such injuries in an effort to reduce injury risk. A proactive approach to injury mitigation could save players days, weeks, or months of time away from play [8]. Predicting injuries, however, is not straightforward and some argue is nearly impossible as current methods are not sufficient to make these predictions in a clinical setting [9–11]. Arnason et al. as well as other researchers have found through logistic regression and further correlation analyses that previous injury is a significant risk factor for future injury [12–18]. Adductor injuries and more broadly lower extremity soft tissue injuries can occur from dynamic and multifactorial changes throughout the course of an athlete's career [17,19–21]. Due to the nature of such injuries, machine learning models have been utilized to account for the complex environments in which an injury occurs. Machine learning models and techniques have been utilized to predict injury among athletes [11,19,22–38]. The predominant machine learning algorithms that have been used to identify injury risk in individuals include logistic and linear regressions, decision trees, random forests (RF), boosting, k-nearest neighbors, artificial neural networks, and support vector machines. Random Forest has been a promising algorithm due to its precision in injury prediction [24,27,30,31,33,38]. RF is able to be trained on smaller sets of data when compared to other techniques such as neural networks. Additionally, RF can account for nonlinear relationships, and while it is considered a black box model that can be difficult to interpret, an understanding of influencing features through variable importance (VIMP) and the directionality of the variables through odds ratios from logistic regression can be obtained [39].

A common problem with injury prediction is the small observation size in the injured class. Techniques such as Synthetic Minority Oversample (SMOTE) [40], and random over and under sampling are generally used to address the issue of data imbalance. In previous injury prediction studies, SMOTE has been utilized to create synthetic injury data to extrapolate from limited data points and has shown promising results in the ability to predict injury [19,24,29,37,41]. However, this is not always the case [25,36]; for example, Carey et al. found in Australian football with a hamstring injury prediction model, using workload variables from wearable sensors such as total distance, running speeds, rate of perceived exertion, and player load, that SMOTE pre-processing showed no such increase in prediction performance defined by the area under the receiver operating characteristic curve (AUC) which explains how well the model was able to separate between the injured and non-injured classes [36]. While SMOTE showed no increase in prediction performance, their injury specific hamstring model resulted in improved model performance compared to their models with inclusion of all non-contact injuries [36]. In contrast, Robles-Palazón et al. assessed model performance with and without class balancing techniques to predict soft tissue, lower extremity injuries in youth soccer players and found multiple techniques such as SMOTE and random undersampling (RUS) increased the AUC in all employed machine learning models [37]. While predictive models for soft tissue lower extremity injury have been constructed, an adductor strain-specific predictive model has yet to be reported in a group of professional soccer players within Major League Soccer (MLS) over multiple seasons.

The objective of this study was to build a soft tissue non-contact adductor strain predictive model utilizing data from a single MLS team over multiple seasons to gain greater insight into the multifactorial relationships best able to predict injury. RF models were constructed and data imbalance was addressed with the utilization of SMOTE. We utilized common input features typically collected at professional soccer clubs such as adductor and abductor strength, inertial measurement unit (IMU) calculated accelerations, decelerations, and change of direction (COD) movements during play, past adductor injury

history, and descriptive factors such as when they played in a game and position. With this information, we aim to gain insight into variables affecting adductor injury risk.

2. Material and Methods

2.1. Study Design and Participants

Participants in this study were all male athletes contracted to a single US MLS team throughout the years 2019–2022. This team was comprised of professional soccer players who are at the highest level of soccer performance in the MLS and within this respective team regarded as the first team. No data was used regarding players in the second, academy, or other organized teams with which players may have been previously contracted, however, if a first team player played in a second team game on occasion, that player's second team game data would be included in this study. Models were structured to satisfy a weekly and monthly predictive future time frame. Weekly data was collected and used to predict an athlete's risk of injury in the following week while monthly data was compiled to predict an athlete's risk of injury in the following month. University of Delaware IRB exemption was given due to the retrospective and deidentified data used within this study. In addition, MLS M-Marc authorization was given to obtain this data and conduct this study.

2.2. Procedures and Instruments

2.2.1. Injury Documentation

Injury logs were reviewed from the years 2019 to 2022. These injury logs were recorded by the athletic trainers and physical therapists employed by the team. Injury logs were created at the time of injury. Information such as injury type, mechanism of injury (MOI), location of injury, field position, and game or practice environment were recorded. For the purposes of this study, we only included soft tissue adductor injuries with non-contact injury mechanisms as this removes the unpredictable nature of external forces acting on the body such as a collision with another player or object. Features from the injury log that were used in this analysis included position type (forward, midfielder, defender, or goalkeeper) and time since adductor injury defined as 0 to 2 months, 12 months or more, and never injured.

2.2.2. Game and Daily Monitoring Systems

Game data was collected through Opta Tracking (Opta Tracking, Stats Perform, London, UK) which tracks which players competed in a game as well as the minutes played for each person in each game. This information was utilized to quantify in-game exposure time for athletes. We included both first and second-team games as many of the contracted first team players would also play in second-team games if they did not have enough in-game exposure time for the week. The second team is the next level below the highest level of professional sport at this club. Opponent and team strategy can lead to different demands between these two levels, however, these differences were not investigated in this study.

Position sensors were worn in a vest with the sensor located between the scapular region (Vector, Catapult, Melbourne, Australia). These sensors utilized both an inertial measurement unit (IMU) and global positioning system (GPS) to measure metrics such as distance run, velocity, acceleration, deceleration, and COD. Athletes wore these sensors during practice and game events between the 2019 to 2022 seasons. Data was stored on the sensors and downloaded to the Catapult platform after practice or game events by the team's Sports Scientist. COD is categorized in the direction of an athlete's dominant and non-dominant side defined by the leg they kick with. Catapult has different acceleration metrics labeled as Inertial Movement Analysis (IMA) acceleration and 2nd generation accel-

eration efforts. The IMA and 2nd generation acceleration efforts are calculated differently within the Catapult system. We found that these metrics were not highly correlated with each other and since both metrics were recorded with Catapult sensors, we included both. However, 2nd generation acceleration efforts had only been recorded from 2020 to 2022 for this cohort and will only be included in select models. In both metrics, thresholds were defined for an acceleration effort. In IMA, this was defined as any qualifying effort over 3.5 G of force and for 2nd generation, any qualifying effort over 3 m/s was included. This was done to avoid any potential misclassification of low/mid acceleration efforts and to assess acceleration efforts in a higher-intensity state. Deceleration efforts were calculated using the IMA corresponding inverse threshold.

Exponential weighted moving averages (EWMA) were calculated for the acceleration, deceleration, and COD efforts and were included as input features to the models. Equation (1) outlines an example of how EWMA is calculated for acceleration efforts.

$$EWMA_{today} = \left(\frac{2}{N+1} \right) \times Accel_{eff,today} + \left(1 - \frac{2}{N+1} \right) \times EWMA_{today-1day} + \dots, \quad (1)$$

This allows greater weight to be given to the most recent collected data point where points taken farther away in time become less influential. This has been found to be a more accurate depiction of the average workload over time rather than a simple averaged effort calculation [42,43]. We constructed this on a weekly and monthly basis for the two model frameworks. EWMA was calculated on a daily basis with regard to the previous 28-day period ($n = 28$) and then those values were averaged for defined week and month time frames.

2.2.3. Strength Assessments

Adductor, abductor, (ForceFrame, VALD Performance, Newstead, Australia) and hamstring (Nordic, NordBord, VALD Performance, Newstead, Australia) strength tests were conducted throughout the 2021 and 2022 seasons. The ForceFrame utilizes uni-axial load cells integrated into pads that can be adjusted for differing isometric assessments and training and has been found to be a reliable and valid testing method for hip strength [44]. Adductor and abductor isometric testing were conducted approximately every eight weeks throughout the 2021 and 2022 seasons. The athlete would be in a supine position with knee flexion at 45 degrees. They were instructed to give a maximal exertion with a 30-second break in between for 3 to 5 repetitions. Athletes were instructed to squeeze onto the pads for adductor strength testing and push for abduction strength testing bilaterally. From this test, we extracted peak force (N) production in the dominant and non-dominant limbs from adduction and abduction as well as interlimb imbalance and ad/abductor imbalance ratio.

Similarly, Nordic utilizes two load cells to measure force and uses a platform to mark the center of knee location to calculate torque with a defined moment arm. This test has shown fair to high inter and intrasession reliability [45,46]. The hamstring Nordic strength assessments were conducted approximately every 6–8 weeks throughout the 2021 and 2022 seasons. Athletes were instructed to kneel on the Nordic platform and keep their hips forward with arms crossed on their chest. Their ankles were secured by hooks in a 90-degree position relative to their lower legs. They were instructed to lean forward under tension for as long as possible before they needed to catch themselves from falling for 1 to 4 repetitions. From these tests, we reported maximum force (N) and interlimb maximum force imbalance.

2.3. Data Analysis

If individuals did not have recorded data from the Catapult sensors, they were logged as having zero acceleration, deceleration, and COD efforts for that day. Efforts were calculated as a summed number of total efforts per week or month depending on the model. For a given participant, strength assessment values were left as missing until the first recorded assessment, after which that value was repeated until a new assessment value was reached. If there were multiple of the same test taken within a given week or month, the values were averaged together to get the respective weekly or monthly strength value.

A Random Forest (RF) algorithm was used to determine if the available set of performance metrics could identify adductor strain injury outcomes. RF was selected due to its ability to robustly iterate through thousands of decision trees to assess the features' ability to predict injury. RF does not need large datasets to train the model in comparison to neural networks for example which may perform better but tend to be more computationally extensive. The utilization of RF also allows us to accurately assess categorical and continuous variables appropriately over time while taking into consideration non-linear and linear relationships that may be present. All features included in the model were determined through assessment of collinearity and selective reduction; namely, pairs of variables with an observed correlation coefficient greater than 0.60 were represented by a single variable in the pair. The single variable was chosen based on its correlation to adductor injury. We set up multiple weekly and monthly models utilizing game, injury, and performance data to assess the model's prediction error and identify which variables had the greatest variable importance (VIMP), a measure that gives a variable's relative importance in producing an accurate prediction; variables with higher VIMPs are more instrumental in producing accurate predictions. Four models were created: Models containing Catapult, game, and injury data from 2019 to 2022 were denoted as Catapult Weekly (CW) and Catapult Monthly (CM). Since consistent strength assessments were only collected in 2021 and 2022 seasons, weekly and monthly models were created with a reduced set of Catapult, game, injury, and strength data and denoted as Catapult Strength Weekly (CSW) and Catapult Strength Monthly (CSM).

SMOTE techniques utilizing a k-nearest neighbors approach to create ten times the synthetic data points from the injured population were employed [40]. We also reduced the non-injured data points via RUS, maintaining a representative group in the process. For every model, in an attempt to avoid too aggressively oversampling, we balanced the data set to ensure that 25% was comprised of injury data points and 75% were comprised of no injury data points. Previous prediction models have employed SMOTE techniques to predict sports injury and mortality rates in differing populations. To balance the data in these studies, research groups have reported upsampling the minority class to 40% of the data set and randomly undersampling the majority class to 60% of the total data set [19,37]. For this analysis, we selected a more conservative 25 to 75% split for minority and majority class sample sizes, respectively.

Initial RFs were run with a defined set of predictor variables (Table 1) through practitioner knowledge of adductor injury and the aforementioned assessment of collinearity. The prediction error was recorded for both the non-injured and injured populations. Two different procedures for testing the model were utilized: (1) Train each model on 2019–2021 data and use 2022 as testing data, and (2) Include all years of data collection, train the model with 80% of the dataset, and test the model on the remaining 20%. In both cases, we then measured variable importance through VIMPs and carried out bootstrapping, a statistical technique featuring repeated re-sampling with replacement, in order to establish the significance of the VIMPs. F1 scores were calculated to assess model accuracy as well as classification error of the injured and non-injured classes.

Table 1. Collected Model Features and Corresponding Definitions.

Position, Injury, and Game Features	
P_{for}	Forward
P_{mid}	Midfielder
P_{def}	Defender
P_{gk}	Goalkeeper
G_{num}	Number of games played
G_{min}	Total minutes played in games
PI_{never}	Record of previous adductor injury
PI_{0-2}	0 to 2 months since previous adductor injury
PI_{12+}	Over 12 months since a previous adductor injury
Adductor, Abductor, and Hamstring Strength Features	
Ab_{maxf}	Abduction maximum force on nondominant leg
Ad_{maxf}	Adduction maximum force on dominant leg
Ham_{maxf}	Hamstring maximum force on dominant leg
Ab_{maximb}	Abduction maximum force interlimb imbalance
Ad_{maximb}	Adduction maximum force interlimb imbalance
Ham_{maximb}	Hamstring maximum force interlimb imbalance
$Ratio_{adab}$	Ratio of adductor over abductor dominant maximum force
IMU/GPS Features	
$Accel_{eff}$	IMA acceleration efforts
$Accel2_{eff}$	2nd generation acceleration efforts
COD_{eff}	COD efforts toward nondominant side
$EWMA_{accel}$	EWMA of acceleration effort
$EWMA_{accel2}$	EWMA of 2nd generation acceleration efforts
$EWMA_{decel}$	EWMA of IMA deceleration efforts
$EWMA_{cod}$	EWMA of COD efforts towards nondominant side

Odds ratios were obtained through logistic regression to assess the directionality (i.e., protective or risk-increasing) of the variables. In some cases, the corresponding logistic regression model did not converge; in these cases, odds ratios are not provided. This phenomenon is generally seen for categorical variables where the crosstabulation of injury and the categorical variable feature some cells having a count of 0. RF model performance is reported through a confusion matrix where a total misclassification percentage is given for injured and non-injured predictions. From this, we also obtain percent error for model-predicted injury and non-injury.

3. Results

From 2019 to 2022, 53 athletes were contracted to the MLS first team. Within this group, eight athletes endured an adductor injury. A total of nine adductor injuries were recorded during this period. Each model had a different number of injuries and athletes recorded due to the variables used and the timing in which the injuries occurred. This information in conjunction with the data points present before and after the utilization of SMOTE and RUS can be seen in Table 2.

Table 2. Model Datasets Pre Balancing (PB) and Post Balancing (POB).

	# Athletes	# Injured Athletes	# Injuries PB	# Injuries POB	# Non-Injuries PB	# Non-Injuries POB
CW (2019–2022)	53	7	8	80	2839	240
CM (2019–2022)	53	6	7	70	733	210
CSW (2021–2022)	37	5	5	50	1165	150
CSM (2021–2022)	37	4	4	40	292	120

Due to the limited sample size of injuries within this cohort, as stated previously, we felt it was necessary to (1) test the model on the 2022 season data as well as (2) utilize the full data set, training on 80% of the data and testing on the remaining 20% which allows the model to be trained on a large subset of the data and tested on unseen data with partial reference to the Pareto principle. Only two adductor injuries were recorded during the 2022 season. The predictive models were not successful in predicting either of these injuries, showing that validation strategy (1) did not demonstrate adequate model predictive performance (i.e., 100% error rate of the injured class). Conversely, validation strategy (2) demonstrated much stronger model performance.

Weekly and monthly structured models were created utilizing player characteristics and performance metrics collected. Results of the RF and logistic regression (LR) of CW and CM models are shown in Table 3; CSW and CSM model results are shown in Table 4.

Table 3. Results of the CW and CM RF and LR. Significant predictors are shaded gray and injury risk % can be computed as 1—odds ratio (OR).

CW Model			
Predictor Variable Name	VIMP	OR	<i>p</i> -Value
Accel _{eff}	0.030	1.011	0.284
COD _{eff}	0.035	1.041	0.289
EWMA _{accel}	0.040	1.082	0.253
EWMA _{cod}	0.050	1.112	0.096
EWMA _{decel}	0.035	1.067	0.274
P _{for}	0.096	4.182	0.152
P _{mid}	0.052	1.118	0.094
P _{def}	0.052	0.176	<i>p</i> < 0.05
P _{gk}	0.007	1.179	0.382
G _{num}	0.037	1.444	0.343
G _{min}	0.034	1.004	0.273
PI _{never}	0.099	0.193	<i>p</i> < 0.05
PI _{0–2}	0.090	28.437	0.317
PI ₁₂₊	0.130	4.288	0.491
CM Model			
Predictor Variable Name	VIMP	OR	<i>p</i> -Value
Accel _{eff}	0.037	1.075	0.155
COD _{eff}	0.028	1.057	<i>p</i> < 0.05
EWMA _{accel}	0.052	1.193	0.160
EWMA _{cod}	0.019	1.083	0.428

Table 3. *Cont.*

CM Model			
Predictor Variable Name	VIMP	OR	<i>p</i> -Value
EWMA _{decel}	0.028	1.073	0.317
P _{for}	0.062	6.363	0.333
P _{mid}	0.011	0.663	0.617
P _{def}	0.009	*	0.198
P _{gk}	0.023	2.029	0.250
G _{num}	0.032	1.422	0.183
G _{min}	0.025	1.002	0.370
PI _{never}	0.110	0.094	<i>p</i> < 0.05
PI _{0–2}	0.144	104.234	0.098
PI ₁₂₊	0.114	7.253	0.144

Note: * signifies the non-convergence of the logistic regression.

Table 4. Results of the CSW and CSM RF and LR. Significant predictors are shaded gray and injury risk % can be computed as 1—odds ratio (OR).

CSW Model			
Predictor Variable Name	VIMP	OR	<i>p</i> -Value
Accel _{eff}	0.016	0.944	0.253
Accel2 _{eff}	0.015	1.001	0.405
COD _{eff}	0.075	1.004	<i>p</i> < 0.05
EWMA _{accel}	0.009	0.923	0.413
EWMA _{accel2}	0.022	1.009	0.091
EWMA _{cod}	0.041	1.141	0.101
EWMA _{decel}	0.019	0.458	0.317
P _{for}	0.227	9.173	<i>p</i> < 0.05
P _{mid}	0.011	0.363	<i>p</i> < 0.05
P _{def}	0.019	0.458	<i>p</i> < 0.05
P _{gk}	0.001	*	0.317
G _{num}	0.006	1.238	0.708
G _{min}	0.011	1.000	0.359
PI _{never}	0.012	0.833	0.230
PI _{0–2}	0.094	14.866	0.228
PI ₁₂₊	0.000	*	1.000
Ab _{maxf}	0.055	1.001	<i>p</i> < 0.05
Ad _{maxf}	0.036	1.002	<i>p</i> < 0.05
Ham _{maxf}	0.043	0.994	<i>p</i> < 0.05
Ab _{maximb}	0.064	1.039	0.183
Ad _{maximb}	0.086	1.014	0.051
Ham _{maximb}	0.033	1.034	0.151
Ratio _{adab}	0.029	0.138	<i>p</i> < 0.05

Table 4. *Cont.*

CSM Model			
Predictor Variable Name	VIMP	OR	p-Value
Accel _{eff}	0.011	0.961	0.359
Accel2 _{eff}	0.053	1.053	$p < 0.05$
COD _{eff}	0.006	1.040	0.391
EWMA _{accel}	0.008	0.970	0.568
EWMA _{accel2}	0.014	1.063	0.484
EWMA _{cod}	0.019	1.101	0.144
EWMA _{decel}	0.008	0.843	0.374
P _{for}	0.186	19.998	$p < 0.05$
P _{mid}	0.005	0.680	0.405
P _{def}	0.005	*	0.096
P _{gk}	0.000	*	1.000
G _{num}	0.045	1.547	0.061
G _{min}	0.019	1.002	0.084
PI _{never}	0.016	0.458	$p < 0.05$
PI ₀₋₂	0.050	9.469	0.389
PI ₁₂₊	0.000	*	1.000
Ab _{maxf}	0.031	1.014	0.500
Ad _{maxf}	0.011	1.006	0.563
Ham _{maxf}	0.022	0.994	$p < 0.05$
Ab _{maximb}	0.061	1.023	0.276
Ad _{maximb}	0.033	1.021	0.509
Ham _{maximb}	0.015	1.006	0.096
Ratio _{adab}	0.037	16.956	$p < 0.05$

Note: * signifies the non-convergence of the logistic regression.

The CW model included 14 variables with position as a defender (P_{def}) and no prior injuries (PI_{never}) identified as significant variables related to decreased odds of adductor injury ($p < 0.05$, Table 3, F1 score (0.942)). Defenders have 82.4% lower odds of adductor injury in the following week compared to those who were not defenders. Individuals who have never had a previous adductor injury have 80.7% lower odds of getting a future adductor strain in the following week as compared to individuals who have had a previous adductor injury.

The CM model included the same 14 variables as CW, however COD efforts (COD_{eff}) and history of ever having a previous adductor injury (PI_{never}) emerged as significant features ($p < 0.05$, Table 3, F1 score (0.932)). Each additional COD effort in a month was associated with a 5.7% increased odds of injury in the following month. Individuals with no previous adductor injury had 90.6% lower odds of enduring a future adductor strain in the following month when compared to those with a previous adductor strain. The logistic regression was not able to converge on the variable, P_{def}, as not all categorical options are present as stated in the statistical approach.

From the CSW, significant variables included P_{for}, P_{mid}, P_{def}, COD_{eff}, Ab_{maxf}, Ad_{maxf}, Ham_{maxf}, and Ratio_{adab} ($p < 0.05$, Table 4, F1 score (0.969)). In the CSW model, forwards had 9.173 times the odds of future adductor injury in the following week compared to non-forwards. Defenders and midfielders exhibited the opposite effect. Defenders had 54.2% and midfielders had 63.7% lower odds of adductor injury in the following week

compared to non-defenders and non-midfielders. Individuals with greater hamstring max force were found to have lower odds (0.6% per 1 N of force) of getting a future adductor strain in the following week. Each unit increase in adductor-to-abductor max force ratio was associated with 86.2% lower odds of adductor injury in the following week. Greater abductor (0.1% per 1 N force) and adductor (0.2% per 1 N force) max force generation was related to greater odds of adductor injury in the following week. Finally, individuals who had more COD efforts had greater odds (0.4% per 1 COD effort) of getting an adductor strain in the following week.

Significant variables identified in the CSM included $Accel2_{eff}$, P_{for} , PI_{never} , Ham_{maxf} , and $Ratio_{adab}$ ($p < 0.05$, Table 4, F1 score (0.962)). In the CSM model, forwards had 19.998 times the odds of a future adductor injury in the following month compared to non-forwards. Hamstring max force showed the same influence as seen in the weekly model, where individuals with greater hamstring max force were found to have lower odds (0.6% per 1 N of force) of getting a future adductor strain in the following month. The ratio of max adductor to max abductor force exhibited differences in findings between the weekly and monthly models. In the monthly model, each unit increase in adductor-to-abductor max force ratio was associated with 16.956 times the odds of a future adductor injury in the following month; this is the opposite direction seen in the weekly model. Those with no previous adductor injury had 54.2% lower odds of future adductor injury in the following month compared to those with a previous injury. Generation 2 acceleration efforts showed increased odds (5.3% per 1 acceleration effort) of adductor injury in the following month.

Between all models, the significant variables that predict adductor injury most often included PI_{never} , $Ratio_{adab}$, Ham_{maxf} , COD_{eff} , P_{for} , P_{def} . Other variables showed significance and should be taken into consideration, however, as seen, results vary between the weekly and monthly predictive frameworks.

Other variables such as cumulative minutes played in a game and number of games played did not emerge as significant influencers in any models. EWMA of acceleration, deceleration, and COD as well as IMA efforts showed no significance. Injuries 0 to 2 months prior and injuries that occurred 12+ months ago did not show significance. The inter-limb imbalance between the adductors, abductors, and hamstrings were not significant predictors of adductor injury.

Overall, we see that the CSW and CSM models had the lowest prediction error for the injured population (Table 5). The CW and CM models also performed well, showing less than 12% misclassification of the injured population. CSM had the lowest injury misclassification of 5% however it did misclassify a small % of the non-injured class as injured. The CSW, however, did not misclassify any non-injured data points and had a 6% error in predicting the injured class. The model that performed best overall was the CSW model with a 1.5% error with the CSM, CW, and CM models following closely behind.

Table 5. Confusion matrix results of all models from RF.

	Error of Injured (%)	Error of Non-Injured (%)	Total Model Error (%)
CW (2019–2022)	8.75	0.83	2.81
CM (2019–2022)	11.43	0.48	3.21
CSW (2021–2022)	6.00	0.00	1.50
CSM (2021–2022)	5.00	0.83	1.88

4. Discussion

Due to the prevalence of adductor injury in the soccer population, we aimed to create a predictive model that identifies characteristics that may put an athlete at higher risk

of injury [1–7]. This study was conducted on a single men's professional team within the MLS utilizing data from the 2019 to 2022 seasons. From this study, we were able to create predictive injury models for the following week and month that were successful in classifying injury and non-injury occurrences within this cohort. We utilized player position, game, Catapult, and strength data to extract features of importance in the models. The findings show that position type, history of previous adductor injury, COD and 2nd generation acceleration efforts, maximum force production of the adductors, abductors, and hamstrings, and the adductor-to-abductor ratio were identified as significant variables in at least one of the four models created. While these findings have the potential for minimal classification error of the healthy population, or false positives, it is important to note that error in classifying the injured population, or false negatives, are present and should be taken into consideration when implementing these findings in practice. Acceptable error in classifying the injured and non-injured class should be determined prior to the application of such models.

The adductor muscle group is a complex group within the core musculature that has multiple functions including adduction, internal rotation, flexion, and extension of the hip as well as joint stability in the frontal plane [47,48]. Due to this, we assumed the model best able to predict injury would take into consideration multiple risk factors associated with injury. We utilized tests that assessed strength in the planes of motion in which the adductors act and assessed quantities of movements where the adductor can be put in environments of great strain such as COD, acceleration, and deceleration movements. Breaking down the findings between each model shows unique variable importance. In the CW model, we found that being a defender and never having a previous adductor injury lowers the odds of future injury in the following week. In the CM model, we observed that never having a previous adductor injury lowers the odds of getting an adductor strain in the following month. As stated in the results, previous injuries 0 to 2 months and 12+ months prior and not show significance which could be due to insufficient reinjury data within the data set. In addition, a greater amount of COD efforts is related to higher odds of injury in the following month.

The CSW model shows midfielders, defenders, greater adductor to abductor max force ratio, and greater hamstring max force generation are related to lower odds of adductor injury in the following week. Forwards, greater COD efforts, greater abduction max force, and greater adduction max force were related to greater odds of adductor injury in the following week. In the CSM model we see no previous adductor injury and greater hamstring max force generation is related to lower odds of adductor injury in the following month whereas forwards, greater generation 2 acceleration efforts, and greater ad/abductor max force ratio are related to higher odds of injury in the following month.

We see forwards tend to have greater odds of injury whereas defenders have lessened odds compared to other position types. This could be explained by the unique demands pertaining to acceleration, top running speeds, distance traveled, and strain that each position type experiences. It has been found that forwards and midfielders tend to have most amount of acceleration efforts on average, and high-intensity distance covered in match play whereas defenders tend to have lower high-intensity distances covered and average number of acceleration movements [49–51]. A team's formation patterns on the field can influence these demands so internal position-type workload comparison may be beneficial. High-intensity movements such as acceleration, deceleration, and change of direction can put the adductors under greater strain and may help explain the differences we see in position type injury risk.

In addition, we found that athletes with no previous history of adductor injury had between 54.2% to 90.6% lower odds of getting injured in the following week or month com-

pared to individuals who had a previous injury. Following muscle injury, structural changes to the musculature and extensive scar tissue can occur which can alter the functionality of the muscle and can lead to reduced strength, greater stiffness, and the potential for reinjury due to the inability to sustain contractile forces [52–59]. This could explain why individuals with no history of adductor injury have lower odds of future injury compared to those with previous injury. The findings from this study provide further evidence that previous injury is a well-defined risk factor of future adductor injury [12–18]. Hamilton et al. highlighted that there are modifiable and non-modifiable factors that could influence one's risk of re-injury [60]. Appropriate rehabilitation can address modifiable factors such as structural issues of the muscle, loss of function, imbalance, and compensatory patterns through other muscle groups. On the other hand, previous injury could be an indicator of non-modifiable intrapersonal characteristics such as position type, age, or hip morphologies that could increase one's risk of injury. It is important to acknowledge and consider non-modifiable factors while addressing modifiable risk factors and appropriately monitoring adductor function for return to sport.

A greater number of COD efforts in game and practice tend to increase the odds of injury whereas a greater Nordic hamstring max force generation has been shown to result in lower odds of adductor injury. When interpreted, we see that the accumulation of COD efforts throughout a week and monthly period can result in much greater odds of injury. As seen in Table 4, for every 1 COD effort, we saw a 0.4% increase in odds of injury in the future week. For example, if player A did 10 COD efforts and player B did 0 efforts for one week, player A would have 4% greater odds of adductor injury in the following week. It is important to note how this can scale when some athletes may experience over 100 efforts in a week. During a COD movement, great strain is put on the adductor musculature as one experiences an adduction movement against an abduction force. These findings support that a greater number of COD efforts is related to adductor injury risk. Consequently, monitoring for greater efforts among individuals for a given week or month could be advised.

The ratio of the adduction-to-abduction max force values showed differing results between the monthly and weekly models. We found that a greater adductor-to-abductor max force ratio would lead to greater odds of injury risk in the following month. However, in the weekly model, individuals with stronger adductor max force compared to abductor max force had lower odds of adductor injury in the following week. The testing windows in this study can introduce repeated measures week-to-week and month-to-month leading to reduced accuracy in acute interpretations. Results suggest that greater adductor max isometric force capability relative to the abductors may have a protective effect on immediate demands experienced by an athlete in the following week, hence reducing risk of injury. However, athletes who maintain such a strength imbalance could be hypothesized to be at a higher risk of adductor injury in subsequent months due to chronic adaptations of greater adductor to abductor max force production. This could influence an athlete's movement strategies and decrease pelvic stability, placing greater strain on the adductors and stabilizing structure during repetitive high force generating movements. Further investigation as to the differences seen in these models and physiological time-dependent responses need to be investigated. There has been supporting evidence found that an increased abductor-to-adductor strength ratio is a risk factor for adductor injury [17,61].

Max hamstring strength was found to be a significant input variable within the prediction models. The functions of the adductors, specifically the adductor magnus, may give context as to why this variable is significant in predicting injury. The adductor magnus is comprised of two sections, the pubofemoral and ischiocondylar portions [62]. Due to the similarities in innervation and actions of the ischiocondylar portion to the hamstrings as a

hip extensor, it brings up the question of whether we can consider hamstring assessments such as the Nordic to be a good assessor for adductor functionality as well. In addition, the Nordic movement relies on stabilization of the trunk and pelvis [63]. While greater all-around strength from the adductors and other pelvic and trunk stabilizers may help sustain the Nordic movement longer, it is not known if it would directly affect max force generation. This would require observation of the muscle activity during this specific movement and potential compensatory patterns that could be present in individuals. However, the adductors play a key role in the stabilization of the pelvis [48,62], which could explain why greater force generation in a functional eccentric movement that relies on pelvic stabilization to assess hamstring strength could be related to lowered odds of adductor injury risk. We found with a 1 N increase in hamstring max force, a person would have 0.6% lower odds of future adductor injury in the following week and month. Since the maximum force is scaled in the hundreds of Newtons, we can see how this has the potential to influence a person's risk of future injury.

In most instances, it is better to have a false negative of the non-injured population if that means having better precision in classifying the injured group. This may lead one to favor the CSM model with a lower % error of the injured class classification even with an error in classifying the non-injured group. When assessing performance between the CW and CM models, the CW has a lower error % of injury classification and overall model misclassification making it more favorable to use. Due to the small sample sizes of the catapult and strength models, the synthetic data was created on a less diverse sample set making this type of model susceptible to overfitting. While the catapult weekly and monthly models also have relatively low injury events, we were able to predict the 70 and 80 injury data points used in these models with low error % despite greater variability amongst the injured class. If we compare weekly and monthly models, the CSW and CSM perform better, however, it is important to take into consideration the increased sample size of the injury class in the CW and CM models.

We have demonstrated the ability to predict adductor strain in an MLS cohort provided that the entire data set is utilized and the influencing variables mentioned are included. When we performed SMOTE on the entire dataset and then proceeded with an 80% train and 20% test split, the random forest model successfully identified a subset of predictive variables for adductor injury in an MLS cohort. As SMOTE interpolates from injury data points to create synthetic injury events, the model would likely not be successful in predicting injuries that had differing characteristics from the original injuries. While the models performed well, it is important to acknowledge the potential for overfitting, however, internal validation through bootstrapping was conducted to obtain *p*-values of variable importance for each predictive input variable. This emphasizes the need for further validation and greater sample sizes when utilizing this information for prediction protocols in a new cohort. Furthermore, with models developed from a subset of the data, we were unsuccessful in predicting new data for a novel season (i.e., 2022 season). This is likely due to the small sample size as we only had two injury occurrences in 2022 and currently alludes to lack of predictive robustness.

As mentioned by Bullock et al., there are shortcomings to the current models that have been utilized to predict sports musculoskeletal injuries [10]. By using the RF and LR methodology, we were able to address continuous and categorical variables appropriately and not assume linearity to give a better representation of the variables assessed.

4.1. Considerations and Future Objectives

While we have demonstrated the utility of RF and SMOTE to predict injury in a single cohort, there are limitations that should be taken into consideration. The number of

non-injury data points recorded within the years of data collection greatly outweighed the number of injured data points collected. Due to this, the utilization of SMOTE techniques to create synthetic instances of injury was needed. When generating synthetic data, it is important to note that instances of future injury could be missed especially if the athlete exhibits different features compared to the current cohort. With the small injury sample size collected during this time frame, application to an injury prevention plan will need further validation and verification.

This model was built from a generalized adductor strain injury set. This means that the adductor strains mentioned could have occurred in any of the six notable adductor muscles. This generalization was needed due to the sample size of the injury. In the future, with a greater injured sample size, it would be advantageous to create predictive models for specific muscles within the adductor group. Each adductor muscle's unique combination of primary and secondary functions as well as location and structure could influence injury risk. In addition, these techniques could be utilized to predict other common injuries in sports. While utilizing performance variables and past injury history is important in efforts to predict injury, further understanding of a person's physiological predispositions, internal load measures, and muscle integrity could provide further information for identifying risk of injury. Soft tissue and non-contact injuries have the potential to be addressed early through training protocols and knowledge of a player's predispositions such as past injury and position type. This information can give practitioners insight into appropriate training and injury mitigation for each of their athletes. For this study, we assessed a single team within the MLS. While this created a more consistent environment based on training and medical protocols, it should be noted that this could influence the model based on this specific controlled environment. For future application, practitioners and researchers should assess model performance in their own club setting.

4.2. Clinical Implications

We have combined RF and SMOTE techniques to identify which common game and performance metrics influence adductor injury risk in a cohort of MLS soccer players. The most common predictors of adductor injury include COD efforts, max hamstring force, ad/abductor max force ratio, position played, and prior injury. Future studies should use a larger sample size, with models applied to unique cohorts and muscle types, prior to the development of injury prevention protocols.

Author Contributions: Conceptualization, R.D., J.S.H., M.H. and B.C.B.; methodology, R.D. and B.C.B.; software, R.D. and B.C.B.; validation, R.D., J.S.H., M.H. and B.C.B.; formal analysis, R.D. and B.C.B.; investigation, R.D.; resources, J.S.H. and B.C.B.; data curation, R.D.; writing—original draft preparation, R.D.; writing—review and editing, R.D., J.S.H., B.C.B. and M.H.; visualization, R.D.; supervision, J.S.H. and M.H.; project administration, R.D. and J.S.H.; funding acquisition, R.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the NSF GRFP and the University of Delaware Mechanical Engineering Helwig Fellowship.

Institutional Review Board Statement: IRB and the MLS M-Marc Medical Committee approval was obtained for this research (code 1874211-1, approval date 19 May 2022).

Informed Consent Statement: Informed consent and approval were obtained from the MLS M-Marc Medical Committee.

Data Availability Statement: Restrictions apply to the availability of these data. Data permissions are needed by the MLS M-Marc Medical committee and are not available for public sharing.

Conflicts of Interest: The contributing author of this paper participated in a fellowship with a professional soccer team to gain sport analytic experience during the 2022 season. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

MLS	Major League Soccer
RF	Random Forest
SMOTE	Synthetic Minority Oversampling
RUS	Random Undersampling
VIMP	Variable Importance
OR	Odds Ratios
AUC	Area under the receiver operating curve
IMU	Inertial Measurement Unit
IMA	Inertial Movement Analysis
GPS	Global Positioning System
COD	Change of Direction
MOI	Mechanism of Injury
EWMA	Exponential Weighted Moving Average
LR	Logistic Regression
PB	Pre Balancing
POB	Post Balancing
CW	Catapult Weekly
CM	Catapult Monthly
CSW	Catapult Strength Weekly
CSM	Catapult Strength Monthly

References

- Ekstrand, J.; Häggglund, M.; Waldén, M. Injury incidence and injury patterns in professional football: The UEFA injury study. *Br. J. Sports Med.* **2011**, *45*, 553–558. [\[CrossRef\]](#)
- Forsythe, B.; Knapik, D.M.; Crawford, M.D.; Diaz, C.C.; Hardin, D.; Gallucci, J.; Silvers-Granelli, H.J.; Mandelbaum, B.R.; Lemak, L.; Putukian, M.; et al. Incidence of Injury for Professional Soccer Players in the United States: A 6-Year Prospective Study of Major League Soccer. *Orthop. J. Sports Med.* **2022**, *10*, 232596712110551. [\[CrossRef\]](#)
- Werner, J.; Häggglund, M.; Ekstrand, J.; Waldén, M. Hip and groin time-loss injuries decreased slightly but injury burden remained constant in men's professional football: The 15-year prospective UEFA Elite Club Injury Study. *Br. J. Sports Med.* **2018**, *53*, 539–546. [\[CrossRef\]](#) [\[PubMed\]](#)
- Werner, J.; Häggglund, M.; Waldén, M.; Ekstrand, J. UEFA injury study: A prospective study of hip and groin injuries in professional football over seven consecutive seasons. *Br. J. Sports Med.* **2009**, *43*, 1036–1040. [\[CrossRef\]](#) [\[PubMed\]](#)
- Lavoie-Gagne, O.; Mehta, N.; Patel, S.; Cohn, M.R.; Forlenza, E.; Nwachukwu, B.U.; Forsythe, B. Adductor Muscle Injuries in UEFA Soccer Athletes: A Matched-Cohort Analysis of Injury Rate, Return to Play, and Player Performance From 2000 to 2015. *Orthop. J. Sports Med.* **2021**, *9*, 232596712110230. [\[CrossRef\]](#)
- Ekstrand, J.; Häggglund, M.; Waldén, M. Epidemiology of Muscle Injuries in Professional Football (Soccer). *Am. J. Sports Med.* **2011**, *39*, 1226–1232. [\[CrossRef\]](#)
- Serner, A.; Weir, A.; Tol, J.L.; Thorborg, K.; Lanzinger, S.; Otten, R.; Hölmich, P. Return to Sport After Criteria-Based Rehabilitation of Acute Adductor Injuries in Male Athletes: A Prospective Cohort Study. *Orthop. J. Sports Med.* **2020**, *8*, 2325967119897247. [\[CrossRef\]](#)
- Drew, M.K.; Raysmith, B.P.; Charlton, P.C. Injuries impair the chance of successful performance by sportspeople: A systematic review. *Br. J. Sports Med.* **2017**, *51*, 1209–1214. [\[CrossRef\]](#) [\[PubMed\]](#)
- Bahr, R. Why screening tests to predict injury do not work—And probably never will. . . : A critical review. *Br. J. Sports Med.* **2016**, *50*, 776–780. [\[CrossRef\]](#)

10. Bullock, G.S.; Mylott, J.; Hughes, T.; Nicholson, K.F.; Riley, R.D.; Collins, G.S. Just How Confident Can We Be in Predicting Sports Injuries? A Systematic Review of the Methodological Conduct and Performance of Existing Musculoskeletal Injury Prediction Models in Sport. *Sports Med.* **2022**, *52*, 2469–2482. [\[CrossRef\]](#)
11. Claudino, J.G.; de Oliveira Capanema, D.; de Souza, T.V.; Serrão, J.C.; Machado Pereira, A.C.; Nassis, G.P. Current Approaches to the Use of Artificial Intelligence for Injury Risk Assessment and Performance Prediction in Team Sports: A Systematic Review. *Sports Med.-Open* **2019**, *5*, 2469–2482. [\[CrossRef\]](#)
12. Arnason, A.; Sigurdsson, S.B.; Gudmundsson, A.; Holme, I.; Engebretsen, L.; Bahr, R. Risk factors for injuries in football. *Am. J. Sports Med.* **2004**, *32*, 5S–16S. [\[CrossRef\]](#)
13. Engebretsen, A.H.; Myklebust, G.; Holme, I.; Engebretsen, L.; Bahr, R. Intrinsic Risk Factors for Groin Injuries among Male Soccer Players. *Am. J. Sports Med.* **2010**, *38*, 2051–2057. [\[CrossRef\]](#)
14. Farrell, S.G.; Hatem, M.; Bharam, S. Acute Adductor Muscle Injury: A Systematic Review on Diagnostic Imaging, Treatment, and Prevention. *Am. J. Sports Med.* **2023**, *51*, 3591–3603. [\[CrossRef\]](#)
15. Langhout, R.; Tak, I.; van Beijsterveldt, A.M.; Ricken, M.; Weir, A.; Barendrecht, M.; Kerkhoffs, G.; Stubbe, J. Risk Factors for Groin Injury and Groin Symptoms in Elite-Level Soccer Players: A Cohort Study in the Dutch Professional Leagues. *J. Orthop. Sports Phys. Ther.* **2018**, *48*, 704–712. [\[CrossRef\]](#)
16. Ryan, J.; DeBurca, N.; McCreesh, K. Risk factors for groin/hip injuries in field-based sports: A systematic review. *Br. J. Sports Med.* **2014**, *48*, 1089–1096. [\[PubMed\]](#)
17. Whittaker, J.L.; Small, C.; Maffey, L.; Emery, C.A. Risk factors for groin injury in sport: An updated systematic review. *Br. J. Sports Med.* **2015**, *49*, 803–809. [\[CrossRef\]](#) [\[PubMed\]](#)
18. Hägg, M.; Waldén, M.; Ekstrand, J. Risk factors for lower extremity muscle injury in professional soccer: The UEFA Injury Study. *Am. J. Sports Med.* **2013**, *41*, 327–335. [\[CrossRef\]](#)
19. López-Valenciano, A.; Ayala, F.; Puerta, J.M.; De Ste Croix, M.B.A.; Vera-Garcia, F.J.; Hernández-Sánchez, S.; Ruiz-Pérez, I.; Myer, G.D. A Preventive Model for Muscle Injuries: A Novel Approach based on Learning Algorithms. *Med. Sci. Sports Exerc.* **2018**, *50*, 915–927. [\[CrossRef\]](#)
20. Meeuwisse, W.H.; 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**, *17*, 215–219. [\[PubMed\]](#)
21. Mendiguchia, J.; Alentorn-Geli, E.; Brughelli, M. Hamstring strain injuries: Are we heading in the right direction? *Br. J. Sports Med.* **2011**, *46*, 81–85. [\[CrossRef\]](#) [\[PubMed\]](#)
22. Leckey, C.; van Dyk, N.; Doherty, C.; Lawlor, A.; Delahunt, E. Machine learning approaches to injury risk prediction in sport: A scoping review with evidence synthesis. *Br. J. Sports Med.* **2025**, *59*, 491–500. [\[CrossRef\]](#) [\[PubMed\]](#)
23. Gabbett, T.J. The development and application of an injury prediction model for noncontact, soft-tissue injuries in elite collision sport athletes. *J. Strength Cond. Res.* **2010**, *24*, 2593–2603. [\[CrossRef\]](#)
24. Huang, Y.; Huang, S.; Wang, Y.; Li, Y.; Gui, Y.; Huang, C. A novel lower extremity non-contact injury risk prediction model based on multimodal fusion and interpretable machine learning. *Front. Physiol.* **2022**, *13*, 937546. [\[CrossRef\]](#)
25. Jauhiainen, S.; Kauppi, J.P.; Krosshaug, T.; Bahr, R.; Bartsch, J.; Äyrämö, S. Predicting ACL Injury Using Machine Learning on Data from an Extensive Screening Test Battery of 880 Female Elite Athletes. *Am. J. Sports Med.* **2022**, *50*, 2917–2924. [\[CrossRef\]](#)
26. Macera, C.A. Predicting lower-extremity injuries among habitual runners. *Arch. Intern. Med.* **1989**, *149*, 2565–2568. [\[CrossRef\]](#)
27. Majumdar, A.; Bakirov, R.; Hodges, D.; Scott, S.; Rees, T. Machine Learning for Understanding and Predicting Injuries in Football. *Sports Med.-Open* **2022**, *8*, 73. [\[CrossRef\]](#) [\[PubMed\]](#)
28. Nassis, G.P.; Verhagen, E.; Brito, J.; Figueiredo, P.; Krstrup, P. A review of machine learning applications in soccer with an emphasis on injury risk. *Biol. Sport* **2023**, *40*, 233–239. [\[CrossRef\]](#)
29. Piłka, T.; Grzelak, B.; Sadurska, A.; Górecki, T.; Dyczkowski, K. Predicting Injuries in Football Based on Data Collected from GPS-Based Wearable Sensors. *Sensors* **2023**, *23*, 1227. [\[CrossRef\]](#)
30. 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**, *10*, 5. [\[CrossRef\]](#)
31. Rossi, A.; Pappalardo, L.; Cintia, P.; Iaia, F.M.; Fernández, J.; Medina, D. Effective injury forecasting in soccer with GPS training data and machine learning. *PLoS ONE* **2018**, *13*, e0201264. [\[CrossRef\]](#) [\[PubMed\]](#)
32. Ruddy, J.D.; Shield, A.J.; Maniar, N.; Williams, M.D.; Duhig, S.; Timmins, R.G.; Hickey, J.; Bourne, M.N.; Opar, D.A. Predictive Modeling of Hamstring Strain Injuries in Elite Australian Footballers. *Med. Sci. Sports Exerc.* **2018**, *50*, 906–914. [\[CrossRef\]](#)
33. Van Eetvelde, H.; Mendonça, L.D.; Ley, C.; Seil, R.; Tischer, T. Machine learning methods in sport injury prediction and prevention: A systematic review. *J. Exp. Orthop.* **2021**, *8*, 27. [\[CrossRef\]](#) [\[PubMed\]](#)
34. Windt, J.; Gabbett, T.J. How do training and competition workloads relate to injury? The workload—Injury aetiology model. *Br. J. Sports Med.* **2016**, *51*, 428–435. [\[CrossRef\]](#)

35. Zelic, I.; Kononenko, I.; Lavrac, N.; Vuga, V. Induction of decision trees and Bayesian classification applied to diagnosis of sport injuries. *J. Med. Syst.* **1997**, *21*, 429–444. [\[CrossRef\]](#)
36. Carey, D.L.; Ong, K.; Whiteley, R.; Crossley, K.M.; Crow, J.; Morris, M.E. Predictive Modelling of Training Loads and Injury in Australian Football. *Int. J. Comp. Sci. Sport* **2018**, *17*, 49–66. [\[CrossRef\]](#)
37. Robles-Palazón, F.J.; Puerta-Callejón, J.M.; Gámez, J.A.; De Ste Croix, M.; Cejudo, A.; Santonja, F.; Sainz de Baranda, P.; Ayala, F. Predicting injury risk using machine learning in male youth soccer players. *Chaos Solitons Fractals* **2023**, *167*, 113079. [\[CrossRef\]](#)
38. Vallance, E.; Sutton-Charani, N.; Imoussaten, A.; Montmain, J.; Perrey, S. Combining Internal- and External-Training-Loads to Predict Non-Contact Injuries in Soccer. *Appl. Sci.* **2020**, *10*, 5261. [\[CrossRef\]](#)
39. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [\[CrossRef\]](#)
40. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. SMOTE: Synthetic Minority Over-sampling Technique. *J. Artif. Intell. Res.* **2002**, *16*, 321–357. [\[CrossRef\]](#)
41. Hassanzadeh, R.; Farhadian, M.; Rafieemehr, H. Hospital mortality prediction in traumatic injuries patients: Comparing different SMOTE-based machine learning algorithms. *BMC Med. Res. Methodol.* **2023**, *23*, 101. [\[CrossRef\]](#)
42. Hunter, J.S. The Exponentially Weighted Moving Average. *J. Qual. Technol.* **1986**, *18*, 203–210. [\[CrossRef\]](#)
43. Murray, N.B.; Gabbett, T.J.; Townshend, A.D.; Blanch, P. Calculating acute: Chronic workload ratios using exponentially weighted moving averages provides a more sensitive indicator of injury likelihood than rolling averages. *Br. J. Sports Med.* **2017**, *51*, 749–754. [\[CrossRef\]](#)
44. Desmyttere, G.; Gaudet, S.; Begon, M. Test-Retest Reliability of a Hip Strength Assessment System in Varsity Soccer Players. *Phys. Ther. Sport* **2019**, *37*, 138–143. [\[CrossRef\]](#)
45. Chavarro-Nieto, C.; Beaven, M.; Gill, N.; Hébert-Losier, K. Reliability of Repeated Nordic Hamstring Strength in Rugby Players Using a Load Cell Device. *Sensors* **2022**, *22*, 9756. [\[CrossRef\]](#)
46. Ogborn, D.I.; Bellemare, A.; Bruinooge, B.; Brown, H.; McRae, S.; Leiter, J. Comparison of Common Methodologies for the Determination of Knee Flexor Muscle Strength. *Int. J. Sports Phys. Ther.* **2021**, *16*, 350–359. [\[CrossRef\]](#) [\[PubMed\]](#)
47. Dostal, W.F.; Soderberg, G.L.; Andrews, J.G. Actions of Hip Muscles. *Phys. Ther.* **1986**, *66*, 351–359. [\[CrossRef\]](#) [\[PubMed\]](#)
48. Neumann, D.A. Kinesiology of the Hip: A Focus on Muscular Actions. *J. Orthop. Sports Phys. Ther.* **2010**, *40*, 82–94. [\[CrossRef\]](#)
49. Rhodes, D.; Valassakis, S.; Bortnik, L.; Eaves, R.; Harper, D.; Alexander, J. The Effect of High-Intensity Accelerations and Decelerations on Match Outcome of an Elite English League Two Football Team. *Int. J. Environ. Res. Public Health* **2021**, *18*, 9913. [\[CrossRef\]](#) [\[PubMed\]](#)
50. Pettersen, S.A.; Brenn, T. Activity Profiles by Position in Youth Elite Soccer Players in Official Matches. *Sports Med. Int. Open* **2019**, *3*, E19–E24. Erratum in *Sports Med. Int. Open* **2019**, *3*, E65.
51. Oliva-Lozano, J.M.; Fortes, V.; Krustrup, P.; Muyor, J.M. Acceleration and sprint profiles of professional male football players in relation to playing position. *PLoS ONE* **2020**, *15*, e0236959. [\[CrossRef\]](#)
52. Sakamoto, M. Effects of Physical Agents on Muscle Healing with a Focus on Animal Model Research. *Phys. Ther. Res.* **2021**, *24*, 1–8. [\[CrossRef\]](#) [\[PubMed\]](#)
53. Järvinen, T.; Järvinen, T.; Kääriäinen, M.; Kalimo, H.; Järvinen, M. Muscle Injuries Biology and Treatment. *Am. J. Sports Med.* **2005**, *33*, 745–764. [\[CrossRef\]](#) [\[PubMed\]](#)
54. Palmer, W.E.; Kuong, S.J.; Elmadbouh, H.M. MR imaging of myotendinous strain. *AJR Am. J. Roentgenol.* **1999**, *173*, 703–709. [\[CrossRef\]](#)
55. Croisier, J.L. Factors associated with recurrent hamstring injuries. *Sports Med.* **2004**, *34*, 681–695. [\[CrossRef\]](#)
56. Wong, S.; Ning, A.; Lee, C.; Feeley, B.T. Return to sport after muscle injury. *Curr. Rev. Musculoskelet. Med.* **2015**, *8*, 168–175. [\[CrossRef\]](#)
57. Kääriäinen, M.; Järvinen, T.; Järvinen, M.; Rantanen, J.; Kalimo, H. Relation between myofibers and connective tissue during muscle injury repair. *Scand. J. Med. Sci. Sports* **2001**, *10*, 332–337. [\[CrossRef\]](#) [\[PubMed\]](#)
58. Bayer, M.L.; Hoegberget-Kalisz, M.; Svensson, R.B.; Hjortshøj, M.H.; Olesen, J.L.; Nybing, J.D.; Boesen, M.; Magnusson, S.P.; Kjaer, M. Chronic Sequelae After Muscle Strain Injuries: Influence of Heavy Resistance Training on Functional and Structural Characteristics in a Randomized Controlled Trial. *Am. J. Sports Med.* **2021**, *49*, 2783–2794. [\[CrossRef\]](#)
59. Fournier-Farley, C.; Lamontagne, M.; Gendron, P.; Gagnon, D.H. Determinants of Return to Play After the Nonoperative Management of Hamstring Injuries in Athletes: A Systematic Review. *Am. J. Sports Med.* **2016**, *44*, 2166–2172. [\[CrossRef\]](#)
60. Hamilton, G.M.; Meeuwisse, W.H.; Emery, C.A.; Steele, R.J.; Shrier, I. Past Injury as a Risk Factor: An Illustrative Example Where Appearances Are Deceiving. *Am. J. Epidemiol.* **2011**, *173*, 941–948. [\[CrossRef\]](#)
61. Tyler, T.F.; Nicholas, S.J.; Campbell, R.J.; McHugh, M.P. The association of hip strength and flexibility with the incidence of adductor muscle strains in professional ice hockey players. *Am. J. Sports Med.* **2001**, *29*, 124–128. [\[CrossRef\]](#) [\[PubMed\]](#)

62. Jeno, S.H.; Launico, M.V.; Schindler, G.S. Anatomy, Bony Pelvis and Lower Limb: Thigh Adductor Magnus Muscle. In *StatPearls*; StatPearls Publishing: Treasure Island, FL, USA, 2025.
63. Narouei, S.; Imai, A.; Akuzawa, H.; Hasebe, K.; Kaneoka, K. Hip and Trunk Muscle Activity during the Nordic Hamstring Exercise. *J. Exerc. Rehabil.* **2018**, *14*, 231–238. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.