



Article

# Predictive Modeling of Injury Risk Based on Body Composition and Selected Physical Fitness Tests for Elite Football Players

Francisco Martins <sup>1,2</sup>, Krzysztof Przednowek <sup>3</sup>, Cíntia França <sup>1,2</sup>, Helder Lopes <sup>1</sup>,  
Marcelo de Maio Nascimento <sup>4</sup>, Hugo Sarmento <sup>5</sup>, Adilson Marques <sup>6,7</sup>, Andreas Ihle <sup>8,9,10</sup>,  
Ricardo Henriques <sup>11</sup> and Élvio Rúbio Gouveia <sup>1,2,9,\*</sup>

- <sup>1</sup> Department of Physical Education and Sport, University of Madeira, 9020-105 Funchal, Portugal
  - <sup>2</sup> Laboratory of Robotics and Engineering Systems, Interactive Technologies Institute, 9020-105 Funchal, Portugal
  - <sup>3</sup> Institute of Physical Culture Sciences, Medical College, University of Rzeszów, 35-959 Rzeszów, Poland
  - <sup>4</sup> Department of Physical Education, Federal University of Vale do São Francisco, Petrolina 56304-917, Brazil
  - <sup>5</sup> University of Coimbra, Research Unit for Sport and Physical Activity (CIDAF), Faculty of Sport Sciences and Physical Education, 3004-504 Coimbra, Portugal
  - <sup>6</sup> CIPER, Faculty of Human Kinetics, University of Lisbon, 1495-751 Lisbon, Portugal
  - <sup>7</sup> ISAMB, Faculty of Medicine, University of Lisbon, 1649-020 Lisbon, Portugal
  - <sup>8</sup> Department of Psychology, University of Geneva, 1205 Geneva, Switzerland
  - <sup>9</sup> Center for the Interdisciplinary Study of Gerontology and Vulnerability, University of Geneva, 1205 Geneva, Switzerland
  - <sup>10</sup> Swiss National Centre of Competence in Research LIVES—Overcoming Vulnerability: Life Course Perspectives, 1015 Lausanne, Switzerland
  - <sup>11</sup> Marítimo da Madeira—Futebol, SAD, 9020-208 Funchal, Portugal
- \* Correspondence: erubiog@staff.uma.pt



**Citation:** Martins, F.; Przednowek, K.; França, C.; Lopes, H.; de Maio Nascimento, M.; Sarmento, H.; Marques, A.; Ihle, A.; Henriques, R.; Gouveia, É.R. Predictive Modeling of Injury Risk Based on Body Composition and Selected Physical Fitness Tests for Elite Football Players. *J. Clin. Med.* **2022**, *11*, 4923. <https://doi.org/10.3390/jcm11164923>

Academic Editors: Jiwu Chen and Yaying Sun

Received: 21 July 2022

Accepted: 18 August 2022

Published: 22 August 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 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) license (<https://creativecommons.org/licenses/by/4.0/>).

**Abstract:** Injuries are one of the most significant issues for elite football players. Consequently, elite football clubs have been consistently interested in having practical, interpretable, and usable models as decision-making support for technical staff. This study aimed to analyze predictive modeling of injury risk based on body composition variables and selected physical fitness tests for elite football players through a sports season. The sample comprised 36 male elite football players who competed in the First Portuguese Soccer League in the 2020/2021 season. The models were calculated based on 22 independent variables that included players' information, body composition, physical fitness, and one dependent variable, the number of injuries per season. In the net elastic analysis, the variables that best predicted injury risk were sectorial positions (defensive and forward), body height, sit-and-reach performance, 1 min number of push-ups, handgrip strength, and 35 m linear speed. This study considered multiple-input single-output regression-type models. The analysis showed that the most accurate model presented in this work generates an error of RMSE = 0.591. Our approach opens a novel perspective for injury prevention and training monitorization. Nevertheless, more studies are needed to identify risk factors associated with injury prediction in elite soccer players, as this is a rising topic that requires several analyses performed in different contexts.

**Keywords:** sports injuries; machine learning; injury prediction; sports monitorization; elite football

## 1. Introduction

Injuries are one of the most significant hampering issues for elite football players [1]. Football is known for its fast-paced and powerful actions [2,3], which might contribute to players' increased risk of injuries [4]. Due to their effects on individuals' mental states and overall teams' performances, elite players' injuries significantly impact the sports business [5,6]. Consequently, elite football clubs have been consistently interested in having practical, interpretable, and usable models as decision-making support for coaches and their technical staff members [7].

From the clinical standpoint, the literature describes the lower limbs as the most affected body zone by sports injuries [4,8–14], particularly for muscle injuries in the thigh area, the quadriceps, and the groin [4,10,15,16]. Since injuries in professional soccer are an increasingly problem, it is crucial that the work done in training sessions reflects the demands of competition, aiming at the development of athletes' performance, which includes injury prevention [17–19].

Machine learning or statistical learning methods are currently tools that can significantly support decision-making in various aspects of the training process. For instance, it has been reported in the literature that some models can optimize training loads [20], which reinforces the applicability of machine learning in improving injury prediction [21,22].

Researchers, managers, and coaches are becoming increasingly involved in injury forecasting, using regular data collection that will allow them to act consciously and intervene on time on this global issue [23]. An investigation conducted over 18 years showed that the total injury rate in practice and competition has dropped during the past years [24]. Although the cause leading to this decrease is still unknown, one potential explanation for this decrease may be related to the effectiveness of injury prevention. If so, it is likely that the motivation of the medical staff at elite football teams is increasing, in terms of implementing and overseeing preventive injury programs [24].

Machine learning offers a modern statistical method that uses algorithms mainly created to deal with unbalanced data sets and enable the modeling of interactions between a large number of variables [25]. In the football context, machine learning has been used in injury prediction, physical performance prediction, training load and monitoring, players' career trajectories, clubs' performance, and match attendance [26].

There has been some research done on elite-football-injury prediction up to this point [23,25,27–31]. In 2019, 96 male elite football players participated in a study throughout a season, with hamstring-strain injuries being the primary anticipated consequence. In that study, the prediction model showed moderate to high accuracy for identifying players at risk of hamstring-strain injuries during pre-season testing [31]. Another example involved 26 elite football players participating in year-long research to forecast non-contact injuries. The authors reported that machine learning was far more accurate than baselines and modern injury-risk-estimating approaches, detecting roughly 80% of injuries with about 50% accuracy [23]. In another study conducted with 132 male elite football and handball players, the prediction model accurately identified elite players at risk of developing muscular injuries [25].

Two types of variables are highlighted in the previous research on predictive modeling of injury risk [30]. The first block of predictor variables is modifiable variables, i.e., training loads or physiological and physical fitness tests. The second type is non-modifiable variables, including demographic variables, anthropometric parameters, and injury histories. Indeed, body composition and physical fitness tests are the most commonly assessed by sports staff given their close relationship with game performance and players' health. Moreover, evaluating and monitoring players' characteristics during the season provides valuable information to understand better players' behavioral changes and support coaches' decision-making in the training and match process. In the sports injury literature, most of the investigation conducted aimed to assess one specific variable at a time to predict injury risk. However, this approach limits the correlation of injury risk and a global interpretation of players' performance in professional football [23]. Therefore, this study aimed to analyze predictive modeling of injury risk based on body composition variables and selected physical fitness tests for elite football players across a sports season.

## 2. Materials and Methods

### 2.1. Participants

Thirty-six players from a professional football team participated in this study. This team competed in the First Portuguese League during the 2020/2021 season.

A description of the variables together with the basic statistics (M—mean value, SD—standard deviation) is given in Table 1. The models were calculated based on 22 independent variables ( $x_1$ – $x_{22}$ ) and one dependent variable ( $y$ ). Independent variables include players' information (sectorial position, age, experience, and number of previous injuries), anthropometric parameters with body composition, and components of physical fitness (flexibility, general strength, explosive strength, speed, agility, and aerobic endurance). The dependent variable is the number of injuries per season. The predictive analysis did not use the data of all athletes. Twenty-four players' data were used. This was due to the fact that some of the athletes were noted to have missing data related to not taking certain physical fitness tests.

**Table 1.** Description of the variables used to construct the predictive model (N = 24).

Variable	Description	M	sd
$x_1$ – $x_3$	Sectorial Position *	-	-
$x_4$	Age (y)	25.45	3.34
$x_5$	Experience (y)	7.29	3.38
$x_6$	Body mass (kg)	80.09	7.07
$x_7$	Height (cm)	182.52	6.01
$x_8$	TBW (L)	51.93	4.66
$x_9$	BFM (kg)	8.2	2.41
$x_{10}$	FFM (kg)	71.2	6.50
$x_{11}$	Previous injury ( $n$ )	1.29	1.63
$x_{12}$	Sit and reach (cm)	34.52	6.79
$x_{13}$	Push-ups ( $n$ )	43.63	8.68
$x_{14}$	Handgrip right (kg)	50.87	9.62
$x_{15}$	Handgrip left (kg)	48.92	8.67
$x_{16}$	CMJ height (cm)	40.14	4.58
$x_{17}$	SJ height (cm)	39.64	4.26
$x_{18}$	LS 5 m (s)	1.16	0.13
$x_{19}$	LS 10 m (s)	1.88	0.16
$x_{20}$	LS 35 m (s)	4.85	0.27
$x_{21}$	Estimated VO2 max (L/kg/min)	50.82	3.98
$x_{22}$	Yoyo (m)	1720	476
$y$	Injury frequency ( $n$ )	0.79	0.72

\*—qualitative variable, M (mean value), sd (standard deviation), Me (median), TBW (total body water), BFM (body fat mass), FFM (fat free mass), CMJ (countermovement jump), SJ (squat jump), LS (linear speed), y (years), kg (kilograms), cm (centimeters), L (liters),  $n$  (number), s (speed), min (minutes), m (meters).

All procedures applied were approved by the Ethics Committee of the Faculty of Human Kinetics, CEIFMH No. 34/2021. The investigation was conducted following the Declaration of Helsinki, and informed consent was obtained from all participants.

## 2.2. Body-Composition Assessment

Body-composition variables were assessed using hand-to-foot bioelectrical impedance analysis (InBody 770, Cerritos, CA, USA). Height was measured to the nearest 0.1 cm using a stadiometer (SECA 213, Hamburg, Germany). The measurements occurred in the early morning, with participants fasting and wearing only their underwear. During the assessment, participants were barefoot, standing with both arms 45° apart from the trunk, with both feet bare on the spots of the platform. A total of 26 evaluations of body composition were considered during the season. Body mass, total body water (TBW), body fat mass (BFM), and fat-free mass (FFM) were retained for analysis.

## 2.3. Physical Fitness Assessment

The sit-and-reach bilateral test was used to evaluate flexibility measurement. A box (32.4 cm high and 53.3 cm long) with a 23 cm heel line mark was used. The participants sat barefoot in front of the box, with both knees fully extended and heels against the box. The

research team held one hand lightly against each participant's knees to ensure complete leg extension. Then, participants placed their hands on top of each other, palms down, and slowly bent forward along the measuring scale. The forward-hold position was repeated twice. The third and final forward stretch was held for three seconds, and the score was recorded to the nearest 0.1 cm.

The push-ups test protocol consisted in performing the highest number of push-ups in one minute, respecting the success criteria judged by the evaluator. The participants started the test in the down position to get correct hand placement and then assumed the up position, from which they did the maximum number of push-ups possible. No cadence was used, although participants were encouraged to execute push-ups with good form but fast enough to obtain the best possible score in a minute. The evaluator independently counted the number of push-ups correctly executed.

The handgrip protocol consisted of three alternated data collection trials for each arm, performed using a hand dynamometer (Jamar Plus+, Chicago, IL, USA). Participants were instructed to hold a dynamometer in one hand, laterally to the trunk with the elbow at a 90° position [32]. From this position, participants were instructed to squeeze as hard as possible, progressively and continuously squeezing the hand dynamometer for about two seconds. The dynamometer could not contact the participant's body; otherwise, the trial was repeated. The best score of the three trials was retained for analysis.

The countermovement jump (CMJ) and the squat jump (SJ) were used to assess lower-body explosive strength [33]. Both protocols included four data collection trials and were performed using the Optojump Next (Microgate, Bolzano, Italy) system of analysis and measurement. In both tests, participants were encouraged to jump to their maximum height. Before data collection, three experimental trials were performed by each participant to ensure correct execution. For the CMJ, participants began in a tall standing position, with feet placed hip-width to shoulder-width apart. Then, participants dropped into the countermovement position to a self-selected depth, followed by a maximal-effort vertical jump. Hands remained on the hips for the entire movement to eliminate any influence of arm swing. If the hands were removed from the hips at any point, or excessive knee flexion was exhibited during the countermovement, the trial was repeated. The participants reset to the starting position after each jump. The SJ protocol testing began with the participant in a squat position at a self-selected depth of approximately 90° of knee flexion, holding this position for the researchers' count of three before jumping. If a dipping movement of the hips was evident, then the trial was repeated. The participants reset to the starting position after each jump.

Linear speed was assessed with maximal sprints at 5, 10, and 35 m, starting from a stationary position. Sprint time was recorded using Witty-Gate photocells (Microgate, Bolzano, Italy). Participants were allowed two trials for each sprinting distance, and the best time was used for analysis.

A yoyo intermittent recovery test was applied to evaluate the athlete's maximum oxygen uptake under repeated high-intensity aerobic exercise [34,35]. The test consists of a 2 × 20 m shuttle run at increasing speeds, interspersed with 10 s of active recovery, controlled by audio signals. The test terminated when the subject was no longer able to maintain the required speed. The total distance and VO<sub>2</sub> maximum record were used as results [36]. The results used were based on the athletes' performance in the yoyo test, which is an indirect method of measuring such variables.

All tests were performed on the same day within a 4 h period in the morning (8 a.m.–12 p.m.). They were conducted by trained staff from the research team, who were familiar with each protocol. All protocols were followed with the utmost rigor, and the organization of the sequence of physical tests was designed to reduce the fatigue factor throughout all tests.

#### 2.4. Injury Report

This study followed the Union of European Football Association (UEFA)'s recommendations for epidemiological investigations. An injury was defined as an event during a scheduled training session or match, resulting in an absence from the next training session or match [37]. Regarding the variables under analysis, the type, zone, and specific location of the injury are complementary variables that identify the part of the body that suffered structural and/or functional changes. The mechanism of injury is intended to understand if the injury was traumatic or if it was contracted by overload. The severity of the injury considers the period, in days, from the athlete's stoppage until resuming field work with the consent of the clinical department. Finally, an injury was marked as recurrent when a player was injured in the same place and type where they were previously affected by an injury. Injury records during the season, including in training and competitive moments, were made daily by the clinical department.

#### 2.5. Predictive Modeling

In this analysis, multiple-input single-output models for prediction were used. The output of the model is a continuous variable and represents the number of occurrences of potential injuries. Therefore, we consider regression-type models, not classifiers. Classic regression models (OLS), shrinkage regression, and stepwise regression were used in the models' calculations. All predictive models were calculated using R Software version 4.2.0 (R Foundation for Statistical Computing, Vienna, Austria, 2022). The implemented methods included:

- The ordinary least squares regression (OLS) used a popular least-squares method, in which weights are calculated by minimizing the sum of the squared errors.
- The Ridge model was calculated using the criterion of performance, which includes a penalty for increased weights. Parameter  $\lambda$  decides the size of the penalty: the greater the value of  $\lambda$  is, the bigger the penalty. The value of lambda can vary from 0 to infinity [38].
- Lasso regression is the model where the mechanism facilitates assigning a penalty to variables, and, in this way, they are eliminated from equations. In Lasso regression [39], the parameter  $s$  (penalty) is used to optimize the model.
- Elastic net (ENET) [40] combines the features of ridge and LASSO regressions. The performance criterion is the so-called naive elastic net. To minimize the criterion, the LARS-EN algorithm was suggested [40], which is based on the LARS algorithm for LASSO regression. In elastic net regression, we have two parameters, penalty  $s$  and  $\lambda$ .
- Stepwise Forward Regression has a forward selection procedure (FS), which begins with an equation that contains only a free expression. The first variable in the equation is the one that has the highest correlation with the output variable. If the coefficient of regression of the variable differs significantly from zero, the variable remains in the equation and another variable is added. The second variable introduced into the equation is the one that has the highest correlation with output, which has been adjusted for the effect of the first variable. If the regression coefficient is statistically significant (using F-test), adding the next variable is implemented in the same way [41,42].

The presented methods were used to calculate models from all variables (Table 1). Additionally, OLS, Ridge, LASSO, and elastic net models have been reimplemented for the best subset of input variables computed from stepwise regression. All models calculated in the study were tested by leave-one-out cross validation (LOOCV). In this method, the data set is divided into two subsets: learning and testing (validation). In LOOCV, the test set is composed of a selected pair of data  $(x_i, y_i)$ , and the number of tests is equal to the number of data  $n$ . During the cross-validation,  $RMSE_{CV}$  error was calculated:

$$RMSE_{CV} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_{-i})^2}$$

where  $n$ —number of patterns,  $y_{-i}$ —the output value of the model built in the  $i$ -th step of cross-validation based on a data set containing no testing pair  $(x_i, y_i)$ ,  $\hat{y}_i$ —the output value of the model built in the  $i$ -th step based on the full data set, and  $RMSE_{CV}$ —root mean square error of prediction.

### 3. Results

Table 2 summarizes the data regarding the participants and injuries characterization of Club Sport Marítimo in the 2020/2021 season. Of the 36 players participating in the study, 23 contracted at least one injury over the 2020/2021 season. Injured players missed an average of 14.3 days per injury. There were 0.9 injuries contracted by the number of participants (34 injuries/36 players) over the study period. Most injuries were classified as traumatic (52.9%). About 50% of the injuries were, according to their severity, moderate, since the athletes missed between 8 and 28 days of training and/or competition. Finally, four of the injuries counted were classified as recurrent.

**Table 2.** Characterization of participants and injuries of CS Marítimo in the 2020/2021 season.

No. of Players	36
No. of Injured Players	23
Total Injuries	34
Average Days Missed Due to Injury	14.3
Injury per Player	0.9
<i>Injury Mechanism</i>	
Traumatic	18 (52.9%)
Overload	16 (47.1%)
<i>Injury Severity *</i>	
Minimal (1–3 days)	4 (11.7%)
Mild (4–7 days)	7 (20.5%)
Moderate	17 (50%)
Severe (+28 days)	6 (17.6%)
<i>Injury Recurrence</i>	
Yes	4 (11.8%)
No	30 (88.2%)

\* Number of days missed by a player due to a sports injury contracted in training or match.

Figures 1–3 summarize the type, area, and specific location of injuries. The lower limbs were the body area most affected by injuries (85.2%). Sprains (35.2%) and muscle injuries (35.2%) were the most recurrent type of injuries throughout the study period, particularly in the ankles (29.4%), quadriceps (11.7%), and hamstrings (11.7%).

Table 3 presents the errors for each model and the sets of predictors calculated by the variable selection methods. The classical OLS regression model has the worst predictive ability, for which the error of  $RMSE = 18.57$ . Such a large error shows that the injury-prediction problem is complex and needs to be regularized by, among other things, using shrinkage regression. The use of shrinkage models (Ridge, LASSO, and elastic net) resulted in a sharp decrease in error and, thus, an improvement in the predictive ability of the model. The best model performing injury-prediction tasks for all predictors is the Ridge model, in which the  $RMSE$  error was 0.698. The optimal Ridge model was calculated for  $\lambda = 82.2$ . Optimizations of all shrinkage models are presented in Figure 4. The LASSO model for all predictors was not calculated because the algorithm does not work properly for such a configuration of the number of variables and patterns. Therefore, the following model used was the elastic net regression model. For elastic net regression, a very small prediction

error was obtained (RMSE = 0.633), and the number of predictors was reduced due to the properties of this method. The result of the elastic net analysis was that the best set of input variables is the set of seven variables:  $x_1$ —sectorial position 1,  $x_3$ —sectorial position 3,  $x_7$ —body height,  $x_{12}$ —sit and reach,  $x_{13}$ — $n$  push-ups,  $x_{15}$ —handgrip (l), and  $x_{20}$ —V35 m.

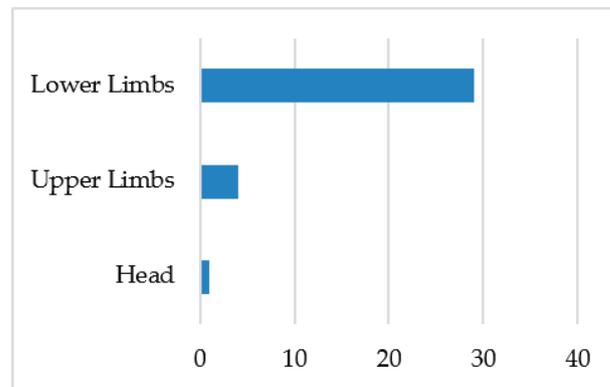


Figure 1. Injury frequency by zone (n).

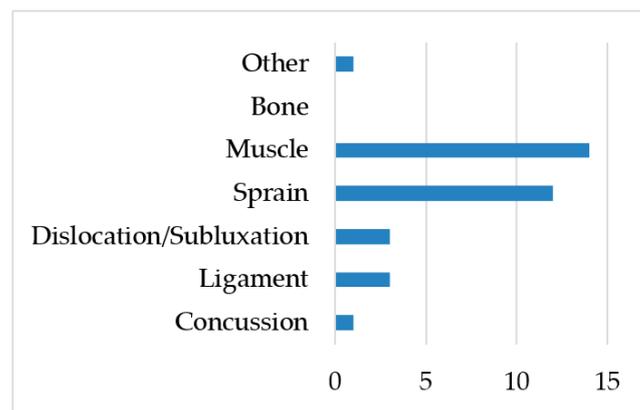


Figure 2. Injury frequency by type (n).

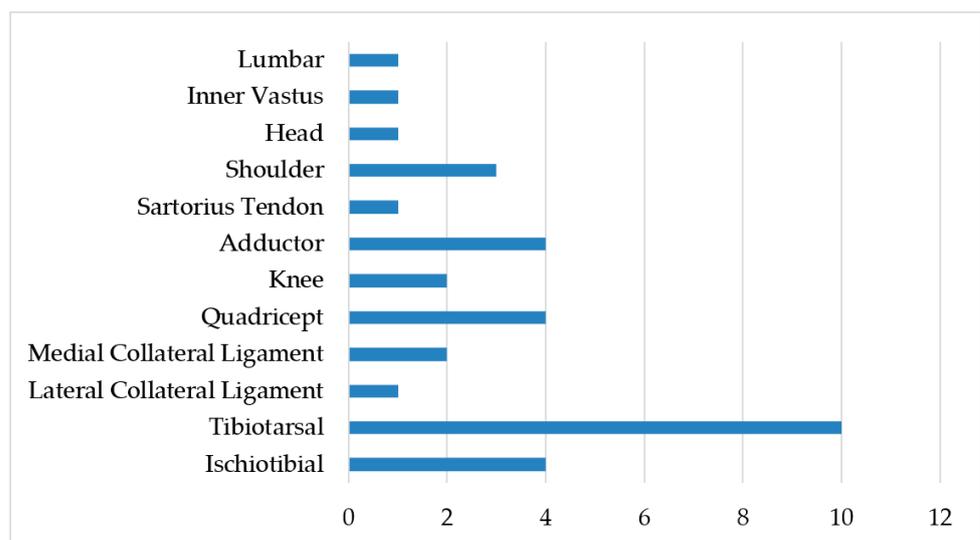
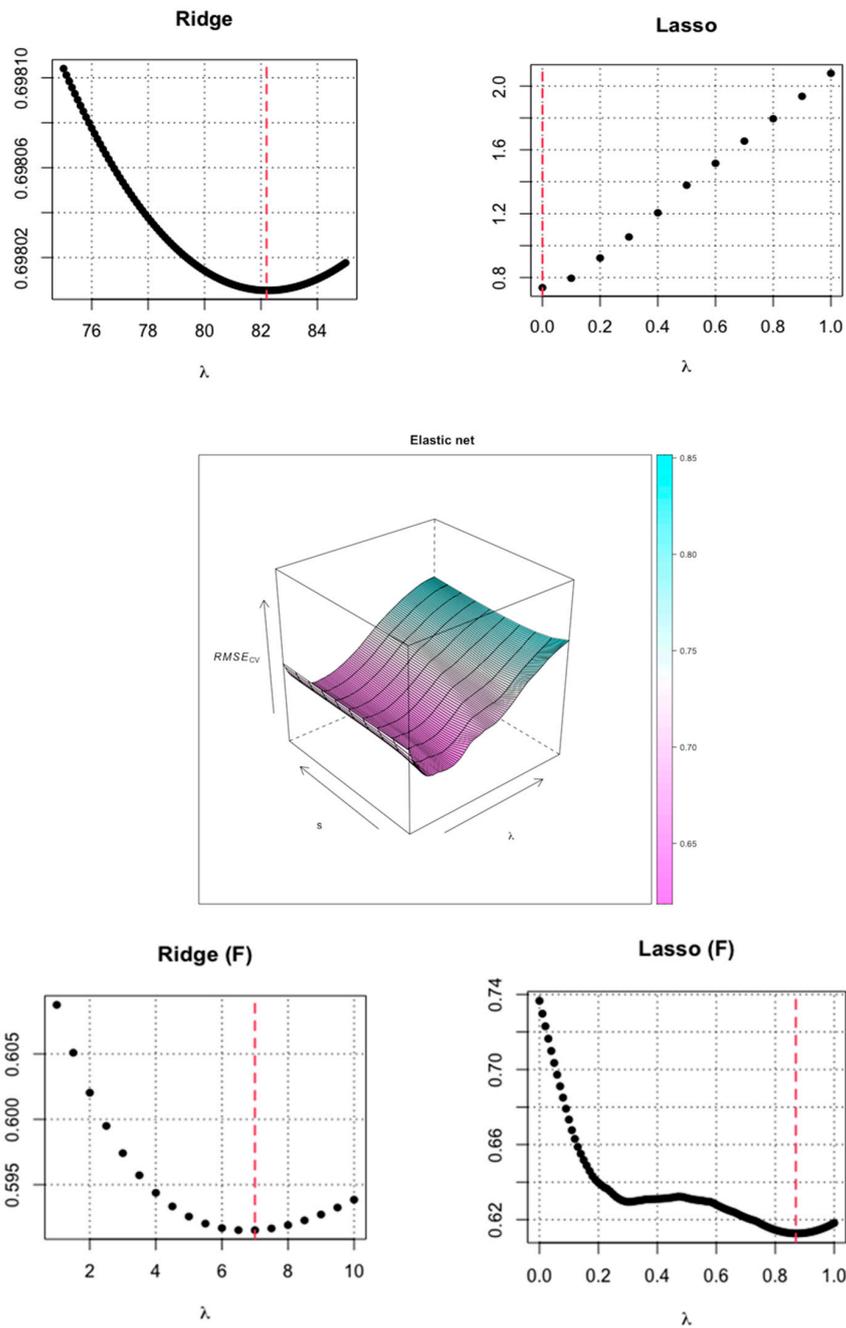


Figure 3. Injury frequency by specific location (n).

**Table 3.** Predictive errors for calculated models.

Method	Predictors	RMSECV	Parameter
OLS	$x_1, x_2, x_3, \dots, x_{23}$	18.57	-
Ridge	$x_1, x_2, x_3, \dots, x_{23}$	0.698	$\lambda = 82.2$
LASSO	$x_1, x_2, x_3, \dots, x_{23}$	0.737	$s = 0$
Elastic net (EN)	$x_1, x_3, x_7, x_{12}, x_{13}, x_{15}, x_{20}$	0.633	$\lambda = 0.1, s = 0.22$
Forward (F)	$x_1, x_{12}, x_{13}, x_{15}$	0.618	-
Ridge (EN)	$x_1, x_3, x_7, x_{12}, x_{13}, x_{15}, x_{20}$	<b>0.592</b>	$\lambda = 17.5$
Ridge (F)	$x_1, x_{12}, x_{13}, x_{15}$	<b>0.591</b>	$\lambda = 7$
LASSO (EN)	$x_1, x_3, x_7, x_{12}, x_{13}, x_{15}, x_{20}$	0.635	$s = 0.55$
LASSO (F)	$x_1, x_{12}, x_{13}, x_{15}$	0.613	$s = 0.87$



**Figure 4.** Cont.

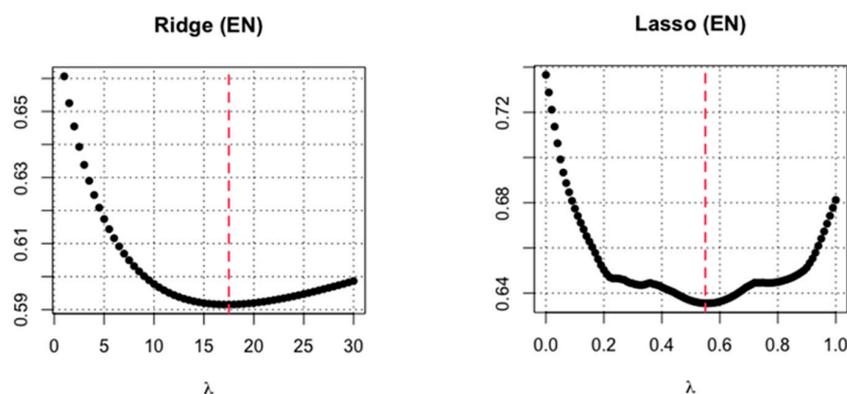


Figure 4. Optimization of predictive models (the red line indicates the optimal model).

The forward regression showed that the significant predictors are  $x_1$  – sectorial position 1,  $x_{12}$ —sit and reach,  $x_{13}$ —n push-ups, and  $x_{15}$ —handgrip I). All the predictors determined by forward regression are contained in the set determined by elastic net regression. The model determined by forward regression generates an error of RMSE = 0.618. The predictors obtained using elastic net (E) and forward regression (F) were used in further predictive analysis. Both sets were used to recalculate the Ridge and LASSO models. The Ridge model with the set calculated by elastic net generates an error of RMSE = 0.592, and a very similar error was obtained for the Ridge model, with the set calculated by forward regression, with RMSE = 0.591. Both Ridge models with new sets of predictors show the best ability. LASSO models for enumerated sets of predictors showed worse predictive abilities than Ridge models. In the case of the best model, the model predicts the number of injury occurrences with an error of 0.59. This means that if a player has three injuries, the model would predict a value from the range of 2.41 to 3.59. The equations for the best models are presented in Table 4.

Table 4. Predictive errors for calculated models.

Method	Equation
Ridge (EN)	$y = 0.01 + 0.10 \oplus x_1 - 0.27 \oplus x_3 + 0.01 \oplus x_7 - 0.01 \oplus x_{12} - 0.01 \oplus x_{13} - 0.03 \oplus x_{15} - 0.45 \oplus x_{20}$
Ridge (F)	$y = -0.28 + 0.35 \oplus x_1 - 0.02 \oplus x_{12} \oplus -0.01 x_{13} + 0.04 \oplus x_{15}$

#### 4. Discussion

This study aimed to analyze predictive modeling of injury risk based on players’ sectorial position, body composition variables (i.e., weight, height, TBW, FAT, and FFM), and selected physical fitness tests, which include sit-and-reach, push-ups, handgrip, CMJ, SJ, 5 m, 10 m, 35 m, and yoyo tests.

This study considered multiple-input single-output regression-type models. It allowed us to select the best model to perform injury prediction tasks, considering all predictors. Previous work on predictive injury risk models is mostly based on classification learning models [31,43,44]. These models’ predictive accuracy ranged from 75% to 82.9% [30]. The present study did not use a categorical variable but rather a continuous variable. A similar solution was presented in another work, where a continuous variable was also placed in the output [45]. A direct comparison of the models’ predictive ability with those presented by other authors is complex because different quality criteria were used.

The value of cross-validation error is important, but a more critical element of the analysis presented was the identification of significant predictors of injury risk. An important part of the analysis was the variable-selection methods, resulting in a very clear and simplified model structure. The simple structure of the model and the linear nature of the methods made it possible to interpret the impact of individual variables on injury

risk. Data-selection mechanisms were also used by other authors who have also used LASSO [44].

According to the data collected for this study, a professional football team can experience 0.9 injuries for every player on the field. This number is noticeably lower than that reported in a study following the analysis of three sports seasons, averaging 1.5 injuries per player [4]. In reality, training load and competitive load—both internal and external—are variables that are related to muscle injuries and that change depending on the situation and level of competition. In this study, sprains and muscular injuries were the most common types of injuries in the lower limbs. The quadriceps and hamstrings were the next most afflicted muscles, followed by the ankles. These results are consistent with the previous findings in the literature [10,12–14,16]. In reality, the lower limbs are under more pressure in this activity because of the tactical–technical maneuvers needed, which justifies their increased risk of damage. Overload injuries were more common than traumatic injuries. A recent investigation also established the existence of such prevalence [4]. In contrast, a different article discovered that overload was the cause of two out of every three injuries in their study [12]. Since there is a strong link between training load and the likelihood of injury, it is imperative to emphasize the significance of appropriately structuring the training cycles according to the players' attributes and physical condition. When individual training loads are measured using the right tools, this process happens more reliably and consistently. Coaches, players, and their technical-support personnel increasingly monitor and evaluate the sports load using a scientific method [46]. In reality, keeping an eye on the training process is essential for assessing the level of athlete weariness, which may help to lower the risk of injury. Soccer involves physical contact and high intensity. Therefore, injury-prevention procedures should take both overload and traumatic injuries into account. Each athlete missed 14.3 days of practice or competition after suffering an injury, on average. This finding differs from that seen in the literature, with players missing an average of seven to eight days owing to injury [4,8,12]. On the other hand, we draw the conclusion that more serious injuries result in a longer period of player absence. This demonstrates the necessity of strengthening all preventative and rehabilitation efforts, while taking into consideration the predictive variables of injury as well as more frequent medical checkups and physical testing. Some authors claim that muscle injuries in soccer are the most common [9,10], converging with our findings. The injury-recurrence rate in our study is consistent with the rates reported in the literature, which range from 8% to 22% [9,47,48]. According to earlier research, these percentage discrepancies may result from the resources available in the individual clinical departments as well as a particular club's infrastructure and material-resource capabilities to respond quickly, in order to maximize the injury prevention and healing process.

Regarding the impact of selected predictors included in the models, first of all, for sectorial position, the defensive and forward sectors were the ones that presented a higher risk of injury. A previous study conducted across three consecutive seasons with 123 Chilean elite male football players also reported that the defensive and forward sectors were the ones that contracted more injuries over the study period [4]. Among 71 Spanish elite male players, forwards were the ones who presented the highest rates in both incidence and severity of injury [14]. Indeed, the literature has described that certain positions, such as fullbacks and forwards, have more demanding tasks both in-game and during training sessions, such as covering greater distances and running with higher intensity than their peers. Overall, fullbacks and forwards perform a total of 29–35 sprints, which is higher than other positions (approx. 17–23 sprints) [49], which may justify their higher injury rates (i.e., hamstring injuries) [50,51]. Therefore, managing training loads appropriately following the physical demands of different sectors and playing positions might be a helpful method to lower the risk of injury in football [52]. Sports agents and coaches should consider load exposure according to players' position, particularly when designing training sessions [52]. Moreover, our results consolidate the need to consider the players' position as a variable to be included in the definition of injury-risk programs.

Another important predictor identified in our study was lower-limb flexibility. The sit-and-reach test is one of the physical fitness tests mostly used to predict the injury risk of elite football players across a sports season. In the literature, several studies have concluded that reduced flexibility in the lower limbs is related to the increased risk of injuries in elite football players [53–56]. Some studies report that it is essential to develop and introduce a standard battery assessment of flexibility in preseason tests, contributing to the awareness of the players' profile [56,57]. The newest Guidelines for Exercise Testing and Prescription from the American College of Sports Medicine reported that maintaining good flexibility in all joints depends on many specific variables, including distensibility of the joint capsule and muscle viscosity, which facilitates movement and may prevent injuries [58]. However, we must acknowledge some limitations on the topic. First, it is not entirely understood if pre-activity stretching unequivocally reduces injuries associated with training load. Secondly, the most recent guidelines recommend direct measures of range of motion (i.e., goniometer and inclinometer) rather than indirect methods, such as sit-and-reach tests assessing flexibility. This means that most of the indirect measures that we most often use in various sports context are coming into disuse. It is recommended that direct measures of range of motion should be used more regularly. In general, the important focus will be that future studies continue to investigate this topic, so we can draw more reliable and valid conclusions regarding the relationship between flexibility and sports injuries.

According to our analyses, the push-up, handgrip, and 35 m linear sprint tests may be reliable predictors of injury risk among elite football players. Besides, height was also one of the variables significantly integrated into injury-prediction models in elite football players. Those variables can be related to each other, since they all end up influencing the players' sports performance. In fact, the main value of this study is directed towards sports monitoring and injury prevention, as we analyzed the relationship between overall strength and height in elite soccer players as predictors of injury, and this is a topic on the rise. In the literature, we identified two studies conducted with youth footballers that have determined that injured players were significantly stronger, bigger, and more experienced than non-injured players [59,60]. This aspect becomes even more relevant when we talk about elite football players, since their demands are higher. The slightest physical differences can make all the difference in the outcome of individual action, dictating the outcome of crucial moments of games and seasons. We believe that these achievements can support future research on the topic to disentangle this complex net of variables that may affect the injury profile.

There are some limitations to this study that need to be acknowledged. The sample size and the fact that we only evaluated the elite players for 26 weeks across 42 weeks of the season are the main limitations of this study. The sample size is related to the number of patterns teaching predictive models. The greater the amount of recorded-injury information is, the better the material for calculating predictive models. Continued collection of learning patterns will improve the predictive ability of the models. Moreover, this is a cross-sectional study, which does not allow a cause-effect of the presented results. However, these results bring important and specific practical implications for those involved in the elite football context, mainly for the topics of injury prevention and training monitorization, since these are issues that are gaining significant attention in the sports business.

## 5. Conclusions

Addressing the need for further studies to identify risk factors for predicting injuries in elite football players, our approach opens a novel perspective on injury prevention and training monitorization, providing a methodology for evaluating and interpreting the complex relations between injury risk and players' performance in elite football. Players' sectorial position, body-composition variables, and physical fitness tests (sit-and-reach, push-up, handgrip, countermovement jump, squat jump, linear speed, and yoyo tests), were all important predictors that may be considered in the injury-risk prevention in elite

football players. It would be an added value if future studies analyzed the influence of body-composition factors and physical fitness tests in elite football teams across different seasons.

**Author Contributions:** Conceptualization, F.M., É.R.G., K.P. and C.F.; methodology, F.M., É.R.G., K.P., C.F. and R.H.; validation, K.P., H.L. and R.H.; formal analysis, F.M., É.R.G., K.P. and C.F.; investigation, F.M., C.F. and R.H.; resources, É.R.G. and H.L.; writing—original draft preparation, F.M., É.R.G., K.P. and C.F.; writing—review and editing, M.d.M.N., H.S., A.M. and A.I.; visualization, M.d.M.N., H.S., A.M. and A.I.; project administration, É.R.G. and H.L.; funding acquisition, É.R.G. and H.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** C.F., F.M. and E.R.G. acknowledge support from LARSyS—the Portuguese national funding agency for science, research, and technology (FCT)—pluriannual funding 2020–2023 (Reference: UIDB/50009/2020). This study is framed in the Marítimo Training Lab Project. The project received funding under application no. M1420-01-0247-FEDER-000033 in the System of Incentives for the Production of Scientific and Technological Knowledge in the Autonomous Region of Madeira—PROCiência 2020.

**Institutional Review Board Statement:** This study was conducted according to the guidelines of the Declaration of Helsinki, was approved by the Ethics Committee of the Faculty of Human Kinetics (CEIFMH N°34/2021), and followed the ethical standards of the Declaration of Helsinki for Medical Research in Humans (2013) and the Oviedo Convention (1997).

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study. Written informed consent has been obtained from all players to publish this paper.

**Data Availability Statement:** The data presented in this study are available upon request from the corresponding author.

**Acknowledgments:** The authors would like to thank all players and their respective legal guardians for participating in this study.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Smpokos, E.; Mourikis, C.; Theos, C.; Manolarakis, G.; Linardakis, M. Injuries and risk factors in professional football players during four consecutive seasons. *Sport Sci. Health* **2021**, *1*–8. [[CrossRef](#)]
2. Konefał, M.; Chmura, P.; Kowalczyk, E.; Figueiredo, A.J.; Sarmiento, H.; Rokita, A.; Chmura, J.; Andrzejewski, M. Modeling of relationships between physical and technical activities and match outcome in elite German soccer players. *J. Sports Med. Phys. Fit.* **2019**, *59*, 752–759. [[CrossRef](#)] [[PubMed](#)]
3. Rice, S.M.; Purcell, R.; De Silva, S.; Mawren, D.; McGorry, P.D.; Parker, A.G. The mental health of elite athletes: A narrative systematic review. *Sports Med.* **2016**, *46*, 1333–1353. [[CrossRef](#)] [[PubMed](#)]
4. Yáñez, S.; Yáñez, C.; Martínez, M.; Núñez, M.; De la Fuente, C. Lesiones deportivas del plantel profesional de fútbol Santiago Wanderers durante las temporadas 2017, 2018 y 2019. *Arch. Soc. Chil. Med. Deporte* **2021**, *66*, 92–103.
5. Häggglund, M.; Waldén, M.; Magnusson, H.; Kristenson, K.; Bengtsson, H.; Ekstrand, J. Injuries affect team performance negatively in professional football: An 11-year follow-up of the UEFA Champions League injury study. *Br. J. Sports Med.* **2013**, *47*, 738–742. [[CrossRef](#)]
6. Hurley, O.A. Impact of player injuries on teams' mental states, and subsequent performances, at the Rugby World Cup 2015. *Front. Psychol.* **2016**, *7*, 807.
7. Kirkendall, D.T.; Dvorak, J. Effective injury prevention in soccer. *Physician Sportsmed.* **2010**, *38*, 147–157. [[CrossRef](#)]
8. Cohen, M.; Abdalla, R.J.; Ejnisman, B.; Amaro, J.T. Lesões ortopédicas no futebol. *Rev. Bras. Ortop.* **1997**, *32*, 940–944.
9. Ekstrand, J.; Häggglund, M.; Waldén, M. Epidemiology of muscle injuries in professional football (soccer). *Am. J. Sports Med.* **2011**, *39*, 1226–1232. [[CrossRef](#)]
10. Hoffman, D.T.; Dwyer, D.B.; Tran, J.; Clifton, P.; Gustin, P.B. Australian Football League injury characteristics differ between matches and training: A longitudinal analysis of changes in the setting, site, and time span from 1997 to 2016. *Orthop. J. Sports Med.* **2019**, *7*, 2325967119837641. [[CrossRef](#)]
11. Lee, I.; Jeong, H.S.; Lee, S.Y. Injury profiles in Korean youth soccer. *Int. J. Environ. Res. Public Health* **2020**, *17*, 5125. [[CrossRef](#)]
12. Noya Salces, J.; Gómez-Carmona, P.M.; Gracia-Marco, L.; Moliner-Urdiales, D.; Sillero-Quintana, M. Epidemiology of injuries in First Division Spanish football. *J. Sports Sci.* **2014**, *32*, 1263–1270. [[CrossRef](#)]
13. Raya-González, J.; de Ste Croix, M.; Read, P.; Castillo, D. A Longitudinal Investigation of muscle injuries in an elite spanish male academy soccer club: A hamstring injuries approach. *Appl. Sci.* **2020**, *10*, 1610. [[CrossRef](#)]

14. Torrontegui-Duarte, M.; Gijon-Nogueron, G.; Perez-Frias, J.C.; Morales-Asencio, J.M.; Luque-Suarez, A. Incidence of injuries among professional football players in Spain during three consecutive seasons: A longitudinal, retrospective study. *Phys. Ther. Sport* **2020**, *41*, 87–93. [[CrossRef](#)]
15. Jones, A.; Jones, G.; Greig, N.; Bower, P.; Brown, J.; Hind, K.; Francis, P. Epidemiology of injury in English Professional Football players: A cohort study. *Phys. Ther. Sport* **2019**, *35*, 18–22. [[CrossRef](#)]
16. Krutsch, W.; Memmel, C.; Alt, V.; Krutsch, V.; Tröß, T.; Meyer, T. Timing return-to-competition: A prospective registration of 45 different types of severe injuries in Germany's highest football league. *Arch. Orthop. Trauma Surg.* **2022**, *142*, 455–463. [[CrossRef](#)]
17. Bradley, P.S.; Noakes, T.D. Match running performance fluctuations in elite soccer: Indicative of fatigue, pacing or situational influences? *J. Sports Sci.* **2013**, *31*, 1627–1638. [[CrossRef](#)]
18. Harper, D.J.; Carling, C.; Kiely, J. High-intensity acceleration and deceleration demands in elite team sports competitive match play: A systematic review and meta-analysis of observational studies. *Sports Med.* **2019**, *49*, 1923–1947. [[CrossRef](#)]
19. Oliva-Lozano, J.M.; Gómez-Carmona, C.D.; Pino-Ortega, J.; Moreno-Pérez, V.; Rodríguez-Pérez, M.A. Match and training high intensity activity-demands profile during a competitive mesocycle in youth elite soccer players. *J. Hum. Kinet.* **2020**, *75*, 195–205. [[CrossRef](#)]
20. Przednowek, K.; Wiktorowicz, K.; Krzeszowski, T.; Iskra, J. A web-oriented expert system for planning hurdles race training programmes. *Neural Comput. Appl.* **2019**, *31*, 7227–7243. [[CrossRef](#)]
21. Stern, B.D.; Hegedus, E.J.; Lai, Y.-C. Injury prediction as a non-linear system. *Phys. Ther. Sport* **2020**, *41*, 43–48. [[CrossRef](#)] [[PubMed](#)]
22. Huang, C.; Jiang, L. Data monitoring and sports injury prediction model based on embedded system and machine learning algorithm. *Microprocess. Microsyst.* **2021**, *81*, 103654. [[CrossRef](#)]
23. 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](#)]
24. Ekstrand, J.; Sprepo, A.; Bengtsson, H.; Bahr, R. Injury rates decreased in men's professional football: An 18-year prospective cohort study of almost 12,000 injuries sustained during 1.8 million hours of play. *Br. J. Sports Med.* **2021**, *55*, 1084–1091. [[CrossRef](#)]
25. López-Valenciano, A.; Ayala, F.; Puerta, J.M.; Croix, M.D.S.; Vera-García, F.; Hernández-Sánchez, S.; Ruiz-Pérez, I.; Myer, G. A preventive model for muscle injuries: A novel approach based on learning algorithms. *Med. Sci. Sports Exerc.* **2018**, *50*, 915. [[CrossRef](#)]
26. Nassis, G.; Stylianides, G.; Verhagen, E.; Brito, J.; Figueiredo, P.; Krstrup, P. A review of machine learning applications in soccer with an emphasis on injury risk. *Biol. Sport* **2022**, *40*, 233–239. [[CrossRef](#)]
27. Brink, M.S.; Visscher, C.; Arends, S.; Zwerver, J.; Post, W.J.; Lemmink, K.A. Monitoring stress and recovery: New insights for the prevention of injuries and illnesses in elite youth soccer players. *Br. J. Sports Med.* **2010**, *44*, 809–815. [[CrossRef](#)]
28. Ehrmann, F.E.; Duncan, C.S.; Sindhusake, D.; Franzsen, W.N.; Greene, D.A. GPS and injury prevention in professional soccer. *J. Strength Cond. Res.* **2016**, *30*, 360–367. [[CrossRef](#)]
29. Venturelli, M.; Schena, F.; Zanolla, L.; Bishop, D. Injury risk factors in young soccer players detected by a multivariate survival model. *J. Sci. Med. Sport* **2011**, *14*, 293–298. [[CrossRef](#)]
30. 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](#)]
31. Ayala, F.; López-Valenciano, A.; Martín, J.A.G.; Croix, M.D.S.; Vera-García, F.; García-Vaquero, M.D.P.; Ruiz-Pérez, I.; Myer, G.D. A preventive model for hamstring injuries in professional soccer: Learning algorithms. *Int. J. Sports Med.* **2019**, *40*, 344–353. [[CrossRef](#)]
32. Gerodimos, V. Reliability of handgrip strength test in basketball players. *J. Hum. Kinet.* **2012**, *31*, 25. [[CrossRef](#)]
33. Bosco, C.; Luhtanen, P.; Komi, P.V. A simple method for measurement of mechanical power in jumping. *Eur. J. Appl. Physiol. Occup. Physiol.* **1983**, *50*, 273–282. [[CrossRef](#)]
34. Schmitz, B.; Pfeifer, C.; Kreitz, K.; Borowski, M.; Faldum, A.; Brand, S.-M. The Yo-Yo intermittent tests: A systematic review and structured compendium of test results. *Front. Physiol.* **2018**, *9*, 870. [[CrossRef](#)]
35. Turner, A.; Walker, S.; Stembridge, M.; Coneyworth, P.; Reed, G.; Birdsey, L.; Barter, P.; Moody, J. A testing battery for the assessment of fitness in soccer players. *Strength Cond. J.* **2011**, *33*, 29–39. [[CrossRef](#)]
36. Bangsbo, J.; Iaia, F.M.; Krstrup, P. The Yo-Yo intermittent recovery test. *Sports Med.* **2008**, *38*, 37–51. [[CrossRef](#)]
37. Häggglund, M.; Waldén, M.; Bahr, R.; Ekstrand, J. Methods for epidemiological study of injuries to professional football players: Developing the UEFA model. *Br. J. Sports Med.* **2005**, *39*, 340–346. [[CrossRef](#)]
38. Hoerl, A.E.; Kennard, R.W. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* **1970**, *12*, 55–67. [[CrossRef](#)]
39. Tibshirani, R. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B* **1996**, *58*, 267–288. [[CrossRef](#)]
40. Zou, H.; Hastie, T. Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B* **2005**, *67*, 301–320. [[CrossRef](#)]
41. Hastie, T.; Tibshirani, R.; Friedman, J. Unsupervised learning. In *The Elements of Statistical Learning*; Springer: Berlin/Heidelberg, Germany, 2009; pp. 485–585.
42. Chatterjee, S.; Hadi, A.S. *Regression Analysis by Example*; John Wiley & Sons: Hoboken, NJ, USA, 2015.
43. Carey, D.L.; Ong, K.-L.; Whiteley, R.; Crossley, K.M.; Crow, J.; Morris, M.E. Predictive modelling of training loads and injury in Australian football. *arXiv* **2017**, arXiv:170604336. [[CrossRef](#)]

44. Rodas, G.; Osaba, L.; Arteta, D.; Pruna, R.; Fernández, D.; Lucia, A. Genomic prediction of tendinopathy risk in elite team sports. *Int. J. Sports Physiol. Perform.* **2019**, *15*, 489–495. [[CrossRef](#)]
45. Ruddy, J.D.; Cormack, S.J.; Whiteley, R.; Williams, M.D.; Timmins, R.G.; Opar, D.A. Modeling the risk of team sport injuries: A narrative review of different statistical approaches. *Front. Physiol.* **2019**, *10*, 829. [[CrossRef](#)]
46. Halson, S. Monitorización de la carga de formación para conocer fatiga en los atletas. *Sports Med.* **2014**, *44*, 139–147. [[CrossRef](#)]
47. Stubbe, J.H.; van Beijsterveldt, A.-M.M.C.; van der Knaap, S.; Stege, J.; Verhagen, E.A.; van Mechelen, W.; Backx, F.J.G. Injuries in professional male soccer players in the Netherlands: A prospective cohort study. *J. Athl. Train.* **2015**, *50*, 211–216. [[CrossRef](#)]
48. Waldén, M.; Häggglund, M.; Ekstrand, J. Injuries in Swedish elite football—A prospective study on injury definitions, risk for injury and injury pattern during 2001. *Scand. J. Med. Sci. Sports* **2005**, *15*, 118–125. [[CrossRef](#)]
49. Di Salvo, V.; Baron, R.; González-Haro, C.; Gormasz, C.; Pigozzi, F.; Bachl, N. Sprinting analysis of elite soccer players during European Champions League and UEFA Cup matches. *J. Sports Sci.* **2010**, *28*, 1489–1494. [[CrossRef](#)]
50. Suarez-Arrones, L.; Torreño, N.; Requena, B.; De Villarreal, E.S.; Casamichana, D.; Barbero-Alvarez, J.C.; Munguía-Izquierdo, D. Match-play activity profile in professional soccer players during official games and the relationship between external and internal load. *J. Sports Med. Phys. Fit.* **2014**, *55*, 1417–1422.
51. Martín-García, A.; Díaz, A.G.; Bradley, P.S.; Morera, F.; Casamichana, D. Quantification of a professional football team’s external load using a microcycle structure. *J. Strength Cond. Res.* **2018**, *32*, 3511–3518. [[CrossRef](#)] [[PubMed](#)]
52. Gabbett, T.J. The training—Injury prevention paradox: Should athletes be training smarter and harder? *Br. J. Sports Med.* **2016**, *50*, 273–280. [[CrossRef](#)] [[PubMed](#)]
53. Hrysomallis, C. Hip adductors’ strength, flexibility, and injury risk. *J. Strength Cond. Res.* **2009**, *23*, 1514–1517. [[CrossRef](#)]
54. 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* (Suppl. S1), 5–16. [[CrossRef](#)]
55. Ekstrand, J.; Gillquist, J. Soccer injuries and their mechanisms: A prospective study. *Med. Sci. Sports Exerc.* **1983**, *15*, 267–270. [[CrossRef](#)]
56. Ibrahim, A.; Murrell, G.; Knapman, P. Adductor strain and hip range of movement in male professional soccer players. *J. Orthop. Surg.* **2007**, *15*, 46–49. [[CrossRef](#)]
57. Witvrouw, E.; Danneels, L.; Asselman, P.; D’Have, T.; Cambier, D. Muscle flexibility as a risk factor for developing muscle injuries in male professional soccer players: A prospective study. *Am. J. Sports Med.* **2003**, *31*, 41–46. [[CrossRef](#)]
58. Liguori, G.; American College of Sports Medicine (ACSM). *ACSM’s Guidelines for Exercise Testing and Prescription*; Lippincott Williams & Wilkins: Philadelphia, PA, USA, 2016.
59. Turbeville, S.D.; Cowan, L.D.; Asal, N.R.; Owen, W.L.; Anderson, M.A. Risk factors for injury in middle school football players. *Am. J. Sports Med.* **2003**, *31*, 276–281. [[CrossRef](#)]
60. Turbeville, S.D.; Cowan, L.D.; Owen, W.L.; Asal, N.R.; Anderson, M.A. Risk factors for injury in high school football players. *Am. J. Sports Med.* **2003**, *31*, 974–980. [[CrossRef](#)]