Title: Predicting injury risk using machine learning in male youth soccer players.

Authors: Robles-Palazón, Francisco Javier[1]; Puerta-Callejón, José M.[2]; Gámez, José A.[2]; De Ste Croix, Mark[3]; Cejudo, Antonio[1]; Santonja, Fernando[4]; Sainz de Baranda, Pilar[1]; Ayala, Francisco[1,3*].

[1]Department of Physical Activity and Sport, Faculty of Sport Sciences, Campus of Excellence Mare Nostrum, University of Murcia. Murcia, Spain.

[2]Department of Computer Systems, University of Castilla-La Mancha, Albacete, Spain.

[3]School of Sport and Exercise, University of Gloucestershire. Gloucester, United Kingdom.

[4]Virgin of the Arrixaca University Hospital, Faculty of Medicine, Campus of Excellence Mare Nostrum, University of Murcia, Murcia, Spain.

*** Corresponding author:** Francisco Ayala. Department of Physical Activity and Sport. University of Murcia. Faculty of Sport Sciences. C/Argentina s/n, 30720. Santiago de la Ribera-San Javier, Murcia (Spain). E-mail address: francisacoayalarodriguez@gmail.com

Predicting injury risk using machine learning in male youth soccer players

Abstract

The aim of this study was twofold: a) to build models using machine learning techniques on data from an extensive screening battery to prospectively predict lower extremity soft tissue (LE-ST) injuries in non-elite male youth soccer players, and b) to compare models´ performance scores (i.e., predictive accuracy) to select the best fit. A sample of 260 male youth soccer players from the academies of five different Spanish non-professional clubs completed the follow-up. Players were engaged in a pre-season assessment that covered several personal characteristics (e.g., anthropometric measures), psychological constructs (e.g., trait-anxiety), and physical fitness and neuromuscular measures (e.g., range of motion [ROM], landing kinematics). Afterwards, all LE-ST injuries were monitored over one competitive season. The predictive ability (i.e., area under the receiver operating characteristic curve [AUC] and F-score) of several screening models was analysed and compared to select the one with the highest scores. A total of 45 LE-ST injuries were recorded over the season. The best fit screening model developed (AUC = 0.700, F-score = 0.380) allowed to successfully identify one in two (True Positive rate = 53.7%) and three in four (True Negative rate = 73.9%) players at high or low risk of suffering a LE-ST injury throughout the in-season phase, respectively, using a subset of six field-based measures (knee medial displacement in the drop jump, asymmetry in the peak vertical ground reaction force during landing, body mass index, asymmetry in the frontal plane projection angle assessed through the tuck jump, asymmetry in the passive hip internal rotation ROM, and ankle dorsiflexion with the knee extended ROM). Given that these measures require little equipment to be recorded and can be employed quickly (approximately 5-10 min) and easily by trained staff in a single player, the model developed might be included in the injury management strategy for youth soccer.

Keywords: screen, associated football, prediction model, adolescent, prevention.

Highlights

- The screening model identified six field-based measures that interact to influence injury risk.
- These measures were based on movement control, asymmetry, range of motion and size features.
- Given that all are modifiable risk factors, at risk players may benefit from interventions.
- The risk factors identified are relatively easy and quick to assess in applied settings.
- The screening model built should be included in the injury management strategy for youth soccer.

1 Introduction

Despite the numerous health-related benefits, the participation in a very physically demanding team sport such as soccer (i.e., associated football) results in a notable increase in injury risk [1]. Epidemiological studies have reported that the frequency and severity of injuries among youth soccer players accelerate and peak during adolescence [2,3], when periods of rapid and non-uniform growth in skeletal structures are experienced, leading to alterations in both physical performance and motor control/function [4,5]. Thigh muscle/tendon strains (hamstring and quadriceps) and knee and ankle ligament sprains and tears (anterior cruciate ligament [ACL] of the knee, anterior inferior tibiofibular ligament of the ankle) are the most commonly diagnosed types of injury in youth soccer players [1,6]. These lower extremity soft tissue (LE-ST) injuries frequently result in players missing sport participation for an extensive period of time [6]. In addition, young players who sustain LE-ST injuries during soccer participation may experience important residual symptoms that can have major negative consequences in their long-term athlete development and limit their ability to engage in exercise and athlete activities later in life [7]. Consequently, soccer-related LE-ST injuries can counter the beneficial health related effects of sport participation at a young age if a child or adolescent is unable to continue participating because of the effects of injury [8].

Most of the LE-ST injuries documented in youth soccer have shown a non-contact mechanism [1] and hence, they might be considered as preventable [9]. Thus, the implementation of multicomponent strategies aimed at mitigating the risk of injury in such cohorts is a big challenge that coaches and physical trainers need to consider. It has been suggested that for an injury prevention measure to be highly effective, its design must be targeted on each player's individual needs [10]. Therefore, the use of a valid field-based screening method that allows coaches and physical trainers to profile injury risk and identify those factors that impact most on the likelihood of sustaining a LE-ST injury in each of their youth soccer players may be a valuable tool to help design tailored preventive measures.

There is a general agreement that LE-ST injury is a multifactorial phenomenon in which several factors of different nature (e.g., personal characteristics, psychological constructs, neuromechanical parameters) might interact among them in a non-linear fashion (complex relationships) and have an impact on the likelihood (i.e., risk) of this one appears (or not) in an athlete (i.e., soccer players) [11–14]. Likewise, epidemiological studies in soccer have documented that the LE-ST injury is an imbalanced phenomenon so that in a typical team the number of players who sustain a LE-ST every competitive season (minority class) is much lower than the non-injured players (majority class) [15]. Most of the screening models available currently to make prospective predictions on new cases of LE-ST injuries in soccer have been built using traditional statistical techniques (mainly binary logistic regression) that were not originally conceived to manage complex (non-linear) and imbalanced phenomena (as the LE-ST is) [16–18]. Furthermore, these models have been designed using information coming from one in isolation or a few factors (no more than six) assessed in a limited sample of soccer players. Consequently, it is not surprising that these traditional models present inadequate performance scores (i.e., predictive accuracy) so that in most of them a clear bias (for many reasons) toward the majority class (known as the negative class) is shown, and therefore, there is a higher misclassification rate for the minority class instances (called the positive examples), which represent the most important concept [19]. In other words, these models usually report high specificity (also called true negative rate [non-injured players who were well-classified]) but very low sensitivity (also called true positive rate [injured players who were well-classified]). Therefore, it has been argued that the complexity of injury means a broader statistical and conceptual approach is needed to make more accurate prospective predictions of new cases of injuries and better understand relationships between risk factors [14,20].

In the last five years, a growing number of studies have used contemporary Machine Learning algorithms (mainly classification [e.g., Random Forest and ADTree] and regression algorithms [e.g., Naïve Bayes and Neural Networks]) which have been specifically designed to deal with imbalanced problems where a large number of factors are involved and resampling methods (e.g., K-fold cross validation, leave-one-out, bootstrapping) to build screening models to profile athletes' injury risk in team sport showing, in most of the cases, promising predictive accuracy [11–13]. Only two recent studies [14,21] have developed screening models using field-based tests to predict injuries through the use of decision tree based classifiers (XGBoost [21] and bagging ensemble method with a J48con decision tree as base classifier [14]) in youth soccer players. In particular, these two studies have built models to classify youth players into two groups, positives (high risk of injury) and negatives (low risk of injury), based on anthropometric (e.g., age, standing and sitting height, body mass), physical fitness (e.g., sprint and jump [vertical and horizontal] performance, agility, lower back and posterior chain flexibility) and neuromuscular (e.g., tuck jump knee valgus angle, unilateral landing peak vertical ground reaction force and asymmetry) measures in elite young male players from the youth academies of six English [14] and seven Belgium [21] premier league soccer clubs, reporting moderate to high levels of sensitivity and specificity, respectively. Furthermore, these studies [14,21] have also identified interactions of asymmetry, knee valgus angle and body size as contributing factors to an injurious profile in elite youth soccer players.

However, it should be acknowledged that a limitation of any prediction model developed through the use of classification algorithms is that its generalisation to individuals with different characteristics (e.g., sport background, exposure to casual factors of injury, physical performance) to those who were employed in its building and validation process may be sub-optimal. In this sense, the well-documented differences in several physical performance measurements [22] between elite and non-elite (i.e., sub-elite or amateur) youth soccer players may lead to a dramatic reduction in the ability of these two currently available screening models to predict LE-ST injuries in the latter cohort. Given that a large proportion of the young participants play for non-professional clubs, engaged in local and regional leagues, and that the injury incidence and severity is still high in this cohort [1], studies aimed at building injury risk factor models to identify non-elite youth soccer players at high risk of LE-ST injury are urgently warranted.

Therefore, the aim of this study was twofold: a) to build models using machine learning techniques on data from an extensive screening battery to prospectively predict LE-ST injuries in non-elite male youth soccer players, and b) to compare models´ performance scores (i.e., predictive accuracy) to select the best fit..

2 Materials and Methods

This study was carried out following the Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD) guidelines [23]. The TRIPOD checklist is provided in online supplementary file 1.

2.1 Participants

A total sample of 301 male youth soccer players from the academies of five different Spanish non-professional soccer clubs were recruited for this study. All players were engaged in regional (non-national) youth soccer leagues of the south-east of Spain. Participants routinely completed from two (most of the weeks in the U11-12 age group) to three (most of the weeks in U13-14, U15-16, and U17-19 age groups) training sessions (90 min each) per week on non-consecutive days and played one competitive match (match duration: U11-12 = 60 min, U13-14 = 70 min, U15-16 = 80 min, U17-19 = 90 min) per week (usually at the weekend) during the season. Participants were included in this study if they met the following criteria: 1) they were free from pain, illness and/or injury during the whole data collection phase and 2) they were regularly involved in soccer training and competition. Players who conveyed the presence of orthopaedic problems that did not allow them to carry out one or more of the field-based tests, or who were transferred to a different club and were not available for follow-up testing at the end of 9-months were excluded. Coaches, parents and children were informed in both oral and written forms, and parental consent to participate in the study was obtained together with assent from participants. Ethical approval was granted by the Ethics and Scientific Committee of the University of Murcia (ID: 1551/2017) in accordance with the Declaration of Helsinki.

Finally, a sample of 260 male youth soccer players of four different age categories (age-based categories [*n*]: U11-12 [78], U13-14 [69], U15-16 [50], U17-19 [63]) completed this study (Table 1). Forty-one players were removed from the initial sample of 301 young based on the exclusion criteria (n = 11 players reported a presence of pain and orthopaedic problems, n = 14 players did not provide the required signed informed consent before the start of the study, and n = 16 players were transferred to another club or left their club before the end of the follow up period).

Table 1. Descriptive anthropometric values (mean ± standard deviation) by age group.

Group	***N***	**Age (years)**	**Body mass (kg)**	**Stature (cm)**	**Leg length (cm)**	**Maturity offset**
U11-12	78	11.1 ± 0.5	39.8 ± 7.4	148.1 ± 6.6	72.8 ± 4.2	-2.4 ± 0.6
U13-14	69	13.3 ± 0.4	51.9 ± 8.6	162.3 ± 7.8	80.8 ± 5.4	-0.7 ± 0.6
U15-16	50	15.0 ± 0.5	62.6 ± 8.5	173.2 ± 6.3	84.9 ± 3.9	1.1 ± 0.6
U17-19	63	17.3 ± 0.8	68.7 ± 8.4	176.6 ± 7.3	86.2 ± 5.5	2.6 ± 0.7

U: under.

2.2 Study design

This study used a prospective cohort design. Particularly, all LE-ST injuries sustained in training and competition during a period of 9 months following the initial assessment session (in-season phase) were tracked for all players. Participants were required to attend their respective club's training facilities during the pre-season phase (September) of the years 2017 (*n* = 175 players) and 2018 (*n* = 85 players) to undergo an assessment of several personal characteristics, psychological constructs, and physical fitness/neuromuscular measures.

2.3 Procedure

The assessment session was split into three different parts. The first part was designed to get data concerning the participants' personal or individual characteristics. Secondly, a number of psychological constructs related to anxiety and mood state were evaluated. Finally, in the third part several physical performance, neuromuscular capability and biomechanical measures were assessed through 10 field-based tests. All measures were taken by six trained and experienced testers (one master and two PhD students and three senior researchers with three, six and more than ten years of experience, respectively), coordinated by the principal investigator (FJR-P) to guarantee standardisation of protocols. All measurements have demonstrated moderate to good reliability (intraclass correlation coefficients [ICCs] > 0.80 and standard error of measurements expressed as percentage [%SEM] < 10%) as it has been described elsewhere [24–29].

2.3.1 Personal or individual characteristics

Personal or individual measures (player position [goalkeeper, defender, midfielder or forward], years of playing soccer, training frequency, dominant leg [determined by the player's preferred kicking leg], self-reported 12 months LE-ST time loss injury history [yes or no], and chronological age) were recorded using an ad hoc questionnaire.

Anthropometric measures (body mass, stature [i.e., standing height], sitting height, body mass index [BMI], and leg and tibia lengths) and maturity status were also measured. Body mass was measured on a calibrated physician scale (SECA 799, Hamburg, Germany). Standing and sitting height were recorded to the nearest 0.1 cm on a measurement platform (SECA 799, Hamburg, Germany) with seated height measured using a box. Leg length was calculated as the length measured in centimetres from the anterior superior iliac spine to the most distal portion of the medial tibial malleolus [25]. Tibia length was defined as the distance between the lateral knee joint line and the lateral malleolus [30]. Stage of maturation was calculated in a non-invasive manner using a regression equation comprising measures of age, body mass, standing height and sitting height [31]. Using this method, maturity offset (calculation of years from peak height velocity [PHV]) was determined (for more information on the personal or individual risk factors recorded, please see online supplementary file 2).

2.3.2 Psychological constructs

The Spanish version of the State-Trait Anxiety Inventory (STAI) questionnaire was used to measure the current state and trait anxiety of the players [32]. This questionnaire consists of 40 items (20 for state and 20 for trait). The state items describe how the athletes feel just at the specific moment when the questionnaire is completed, whereas the trait items describe the athletes' general anxiety level. For the purposes of this research, only the trait anxiety was analysed.

Mood states were evaluated using the Spanish adapted version for adolescent athletes of the Profile of Mood States (POMS) scale [33]. This version comprises seven different psychological factors (tension, depression, anger, vigour, fatigue, confusion, and friendliness) in a 33-item scale.

The Spanish version of the Psychological Characteristics Related to Sport Performance questionnaire (CPRD) was used to measure the following psychological characteristics: stress control, influence of performance evaluation, motivation, team cohesion and mental skills [34]. The questionnaire consists of 55 items graded in a 5-option Likert scale (from totally disagree to totally agree) (for more information on psychological risk factors recorded, please see online supplementary file 3).

2.3.3 Physical fitness, neuromuscular capability and biomechanical measures

Players completed a standardised dynamic warm-up, which included whole body exercises emphasising dynamic mobilisation and gradually progressing in intensity [35], before the physical performance, neuromuscular capability and biomechanical measures were taken. In particular, these measures were concurrently recorded using a randomised circuit style approach (due to time constraints) (Figure 1) from six jump tests, a linear 30 m sprint test, the ROM-Sport battery, Y-Balance test and Illinois agility test.

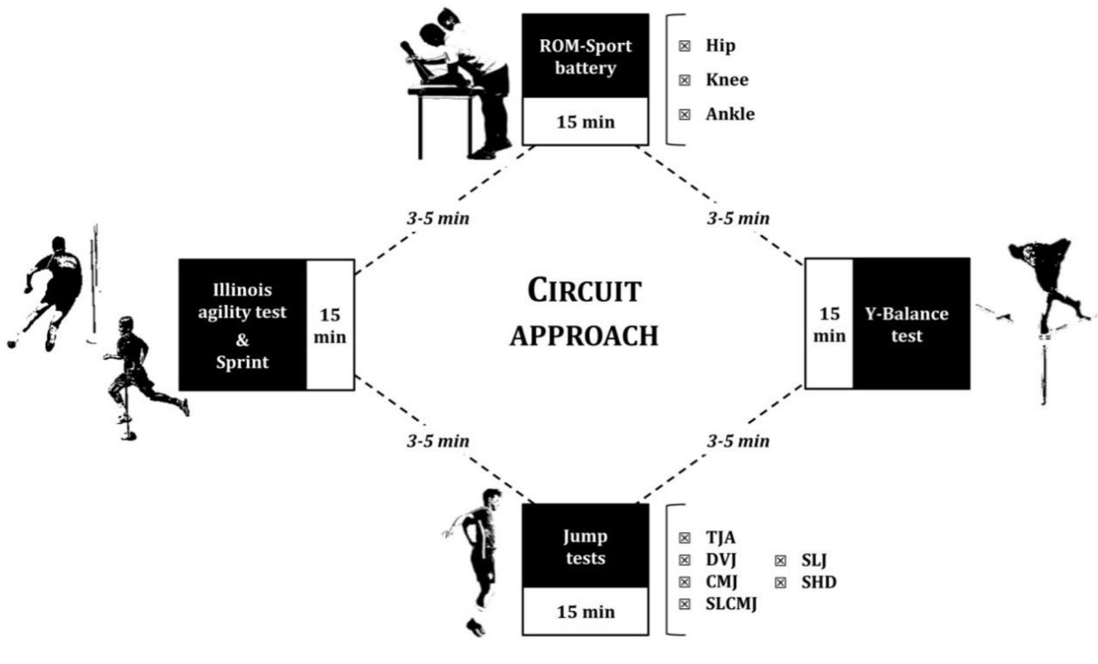

Figure 1. Circuit style approach.

Jump tests

Four vertical and two horizontal jump tests were performed and several measures of performance, kinematic and kinetic variables and neuromuscular parameters were extracted from them. Three to five attempts of each jump test were performed. For each variable, the best absolute score recorded in the attempts carried out was selected for the subsequent analysis (for more information on measures obtained from the Jump tests, please see online supplementary file 4).

Vertical jump tests:

Tuck jump assessment (TJA). Tuck jumps were performed in place for 10 consecutive seconds following the procedure previously suggested by Myer et al. [36]. Each participant's technique was assessed at frontal and sagittal planes. A 2-dimensional video cameras (model: Lumix DMC-FZ200; Panasonic, Japan) were positioned in both planes at a height of 0.70 m and a distance of 5 m from the landing area to capture the test and grade each player's technique retrospectively. Afterwards, frontal plane projection angles (FPPA) at the point of maximum knee flexion were analysed, and the presence of knee valgus was subjectively classified as minor (<10°), moderate (10°–20°) or severe (>20°) following the methodology described by Read et al. [27]. Additionally, hip flexion (HF), knee flexion (KF), and ankle flexion (AF) was assessed at initial contact and peak maximum flexion in the sagittal plane [29]. All scores were marked by two experienced testers in 2-D landing kinematic assessments.

Drop vertical jump (DVJ). A double leg drop vertical jump from a box height of 40 cm and without arm swing was performed on a contact platform connected to the Ergo tester (Ergo Jump Bosco System, Italia) unit [37]. Both jump height and reactive strength index (RSI = jump height/contact time) were considered to assess stretch-shortening cycle (SSC) function and hence, recorded. A 2-dimensional landing kinematic analysis following the methodology described for the TJA was also carried out. In addition to the FPPA, the knee medial displacement (KMD) (expressed as the displacement measure [d2–d1] between the initial contact [d1] and the maximal peak knee flexion [d2]) [30] the knee-to-ankle separation ratio (KASR) (defined as the ratio of distance between knees and ankles during peak knee flexion [KASR = knee/ankle]) [28] and the knee separation distance

(KSD) (expressed as the difference [d2-d1] between knee separation distance at the initial contact [d1] and the peak knee flexion [d2]) [28] were also used to assess knee valgus during DVJ tests. All trials were retrospectively analysed by the same two experienced testers in 2-D landing kinematics assessments.

Countermovement jump (CMJ). A double leg countermovement jump without arm swing was performed on a contact platform connected to the Ergo tester (Ergo Jump Bosco System, Italia) unit. Jump height was recorded for subsequent analyses.

Single leg countermovement jump (SLCMJ). A single leg (dominant and non-dominant) countermovement jump was also performed on a force platform (9286AA, Kistler, Switzerland). Height, peak vertical ground reaction force (pVGRF) during take-off and landing, and peak landing force timing (pLFT) were captured at a sampling rate of 1000 Hz. A threshold of >10 N to determine contact and <10 N to determine flight moments was used, and no filter was applied to the data obtained for subsequent analyses [38]. The pVGRF at take-off and landing were normalised to body weight (BW), and side-to-side differences for each of these variables were calculated. Asymmetries in all SLCMJ variables were determined when bilateral differences were ≥ 10%.

Horizontal jump tests:

Standing long jump (SLJ). Jump distance in a SLJ was measured to the nearest centimetre from the starting line to the player’s heel with a standard tape measure. Free movement of the arms was allowed during the test.

Single hop for distance (SHD). Jump performance in a SHD was also measured for dominant and non-dominant legs [39]. The jump distance in cm was then normalised and presented as percentage of leg length (SHD/leg length*100 = %leg length). Bilateral differences were calculated and asymmetry was considered when differences ≥ 10%.

Sprint

Time during a 10-20 and 30 m sprint in a straight line was measured by means of three pairs of Microgate Witty photocells (Microgate, Italy) placed 1.0 m above the ground level. Each sprint was initiated from an individually chosen standing position, 50 cm behind the photocell gate, which started a digital timer. The theoretical maximal force (F0), velocity (V0), maximal power output (Pmax) and mechanical effectiveness of ground force application (ratio of force [RF] and decrease in the RF over acceleration [DRF]) during a 30m-sprint were also analysed. For this purpose, all sprint trials were recorded through an iPad Air (Apple Inc., USA) and retrospectively analysed by a single tester using the *MySprint* app [26]. The analysis of sprint force-velocity profile in youth athletes has proven to be reliable in previous research [40] (for more information on measures obtained from the Sprint, please see online supplementary file 5).

ROM-Sport battery

The passive hip extension (PHE), hip adduction with hip flexed 90° ($PHAD_{HF90°}$), hip flexion with knee flexed (PHF_{KF}) and extended (PHF_{KE}), hip abduction with hip neutral (PHABD) and hip flexed 90° ($PHABD_{HF90°}$), hip external (PHER) and internal (PHIR) rotation, knee flexion (PKF), ankle dorsiflexion with knee flexed (ADF_{KF}) and extended (ADF_{KE}) ROM measures of the dominant and non-dominant legs were evaluated according to the methodology suggested by Cejudo et al. [24]. For each joint ROM measure, side-to-side differences were also calculated. When a side-to-side difference ≥ 8° was found, players were categorised as showing bilateral asymmetries [41] (for more information on data collected with the ROM-Sport battery, please see online supplementary file 6).

Y-Balance test

Dynamic postural control was evaluated using the Y-Balance test [25]. The distance obtained in each direction (anterior, posteromedial, and posterolateral) was normalised by dividing by the previously measured leg length to standardise the reach distance ([excursion distance/leg length] x 100 = % leg length) [25]. Bilateral differences between dominant and non-dominant legs were also calculated for each distance, and differences ≥ 10% for anterior, posteromedial, and posterolateral directions were considered as asymmetries. Finally, to obtain a global measure of the balance test for each leg, data from each direction were averaged to calculate a composite score (for more information on measures obtained from the Y-Balance test, please see online supplementary file 7).

Illinois agility test

Players' agility was assessed using the Illinois agility test, which has been commonly used in measuring agility in soccer [42]. The length of the zone was 10 m, while the width (distance between the start and finish points) was 5 m. Four cones were placed in the centre of the testing area at a distance of 3.3 m from one another. Four cones were used to mark the start, finish, and two turning points. The participants started the test lying face down, with their hands at shoulder level. The trial started on the "go" command, and the participants began to run as fast as possible. The trial was completed when the players crossed the finish line without having knocked any cones over. Time was measured using a photocell system (Microgate Witty photocells; Microgate, Italy).

2.3.4 Injury surveillance

The procedures for data collection and reporting injury occurrences described in the International Consensus Statement were followed in the current research [43]. For the purpose of this research, an injury was defined as any non-contact, soft tissue (muscle, tendon, and/or ligament) injury sustained by a player during a training session or competition which resulted in a player being unable to take a full part in future soccer training or match play (time loss injury definition). Injuries were classified as non-contact where no clear contact or collision with another player, object or ball occurred. Only lower extremity injuries were considered for the analysis as these incidents are the most common at youth soccer practice [1]. All injuries were recorded by team doctors and physiotherapists of each club, and players were considered injured until the medical staff allowed them to fully participate in training and competition. Injury severity was defined as slight/minimal (1-3 days), minor/mild (4-7 days), moderate (8-28 days), and severe (>28 days) based on lay-off time from soccer.

The club medical staff documented LE-ST injuries on an injury report form described elsewhere [43]. As some inconsistencies in the diagnosis of minimal LE-ST injuries by medical staff teams were found at the end of the 9-month follow-up period, only LE-ST injuries showing a time loss of ≥ 4 days were chosen for the subsequent statistical analysis. Due to the confounding effects of previous injuries, only the first occurring injury for each player during the season was considered in the analyses [14,21].

2.4 Statistical analysis

Data from questionnaires and field-based tests were collected in paper format and transferred into a spreadsheet using a double manual data entry processing technique [44]. Identified discrepancies were corrected upon agreement to reach an error level of 0%. After having performed a rigorous data cleaning process (identified anomalies or errors were corrected [32 cases]) we had an imbalanced (displaying an imbalance ratio of 0.21) and a high-dimensional data set comprising of 260 male youth soccer players and 135 potential risk factors. In this research, an anomaly or error was defined as a value or score that could not be classified as true or real because of the consequence of a human error or a machine failure. An example of an error was a jumping height value of 256 cm since it is impossible for an adolescent to jump as such height.

To assess the performance of the algorithms selected, the fivefold stratified cross-validation technique was applied. The fivefold stratified cross validation was repeated a hundred times and results were averaged over the runs to obtain a more reliable estimate for the predictive ability. A wide variety of classification performance measures may be obtained from the stratified cross-validation technique. A well-known approach to produce an evaluation criterion is to use the receiver operating characteristic (ROC) curve. Thus, the area under the ROC curve (AUC) was employed as a measure of a classifier's performance for evaluating which models showed high (0.90–1.00), moderate (0.70-0.90), low (0.50-0.70) and fail (<50) scores [45]. For the purpose of this study, only algorithms with performance scores (AUC) above 0.70 were considered acceptable. Also, two extra measures from the confusion matrix were selected as evaluation criteria: true positive (TP) rate and true negative (TN) rate. In imbalanced domains, when the AUC has reached a high score (> 0.70), the classification performance may not be as good as the AUC value reflects because plenty of "clear" negative samples (instances that can be clearly classified into the negative label of the class variable) exist in the dataset. These clear negative samples may increase the AUC score, but a few other "border line" negative samples remain mixed with the positive samples (i.e., class overlapping and/or small disjuncts), which are difficult to distinguish and classify by the algorithms. These few remaining border line negative samples may decrease performance (when some of them are wrongly classified [i.e., false positive]), including precision and recall, while very slightly influencing the AUC score. In consequence, Zou et al. [46] recommend using the F-score together with the AUC as a classification measurement for imbalanced problems.

Similar to previously published studies aimed at building prediction models to identify elite soccer [11,13] and futsal [12] players at high (or low) risk of injury based on a supervised learning perspective (i.e., it is defined by its use of labeled datasets according to the class variable [injury yes vs. injury no] to train algorithms that classify data or predict outcomes), the taxonomy for external (resampling techniques), internal (ensemble techniques) and cost-sensitive methods for learning with imbalanced data sets suggested by López et al. [19] and Elkarami et al. [47] was applied. A brief description of each of the techniques employed is provided in online supplementary file 8 as well as in previous studies [12,13]. According to Robertson [48] four different subsets or categories of base learning algorithms can be defined according to their internal functioning to help sports practitioners improve their decision-making processes on training prescription to optimise sports performance and mitigate injury risk: a) regression algorithms (estimating relationships between variables on a continuous scale [e.g., linear regression, neural networks]), clustering algorithms (grouping sets of items based on their levels of similarity to one another [e.g., K-means and hierarchical]), rule-based algorithms (extracting rules from data based on frequency and predictability [e.g., support vector machines and decision rules]) and classification algorithms (identifying which category an instance belongs to and base on a training set of data [e.g., decision trees and Random Forest]). Therefore, six well-known learning algorithms (C4.5, ADTree, SMO, KNN, and Random Forest [RF]) from the categories established by Robertson [48] were selected as base classifiers to be used in the resampling, ensemble, and cost sensitive methods. With all algorithms applied to all base classifiers, a total of 72 models were generated. To allow comparison of the constructed models to a baseline model, a ZeroR classifier was also used.

Some specific pre-processing tasks (missing data imputation and feature selection) were exclusively carried out in the training folds so that the classification task could be performed appropriately. In particular, missing data were substituted by the mean value of the corresponding variable according to the age category of the players.

Due to the high dimensionality of the data set, before running the algorithms included in the taxonomy described in online supplementary file 8, a feature selection process was conducted with the aim of helping base classifiers to reduce the feature space and eliminate irrelevant, weakly relevant and/or redundant features. Particularly, the metaclassifier "attribute selected classifier" available in Weka's repository was employed. We used as attribute evaluator the classify subset evaluator filter [49] and the GreedyStepwise as search technique. To interpret and visualise the behaviour and relevance of the variables selected, the Shapley Additive exPlanations (SHAP) approach (SHAP summary plot) was used [50]. This approach visualises every single player or injury case and gives an overview of the variables in the model by order of importance (vertically listed features), with the top ones having a higher global impact on the model than bottom ones. The SHAP-values represent the impact of a variable in the decision-making process. Dots representing the SHAP-values for each feature value of a player in the dataset are plotted horizontally next to the feature. Negative SHAP-values represent a higher probability of a positive prediction (i.e., being injured). Each dot is coloured by the value (i.e., measured value) of the feature for an individual.

3 Results

3.1 Lower extremity soft-tissue injuries epidemiology

There were 61 LE-ST injuries over the 9-month follow-up period. Of them, 36 were classified as thigh muscle (18 hamstrings, 8 quadriceps, and 10 adductors) injuries, 9 as knee (5 ligament sprains) injuries, and 7 as ankle (all ligament sprains) injuries. The distribution of injuries between legs was 43 dominant leg and 18 non-dominant leg. A total of 26 injuries happened during training sessions and 35 during matches. With regard to severity, most injuries were categorised as moderate (*n* = 40), while only 6 incidents were classified as severe injuries and 15 as minor/mild injuries. Thirteen players sustained multiple LE-ST injuries during the observation period (10 players were injured twice and three players three times) and thus, only their first incident (i.e., the index injury) was used for the analyses. Consequently, 45 LE-ST injuries were finally used to build the prediction models.

3.2 Prediction models for lower extremity soft tissue injuries

As all the algorithms employed in this study can be found in the Weka experimenter, only the scheme (and not the full code) of the algorithm finally selected is displayed in online supplementary file 9 and the model is publicly available on https://data.mendeley.com/datasets/2mw6w556yg/1 in order to allow practitioners to use it with their male youth soccer players.

The feature selection process conducted in the data set identified a subset of six measures as the most relevant (considering the individual predictive ability of each feature as well as the degree of redundancy among them) (Table 2) on which was subsequently applied the taxonomy of learning algorithms explained in the "Materials and Methods" section.

Table 2. Features selected after having applied the classify subset evaluator filter to the data set.

Name	**Labels**
KMD (dominant leg) [DVJ]	0 (varus), 1 (slight valgus), 2 (moderate valgus) or 3 (severe valgus)
BMI	Numeric
ROM-ADF$_{KE}$ (dominant leg)	Numeric
Landing BIL-pVGRF [SLCMJ]	0 (Asymmetry) or 1 (No Asymmetry)
ROM-BIL-PHIR	0 (Asymmetry) or 1 (No Asymmetry)
BIL-FPPA [TJA]	0 (Asymmetry) or 1 (No Asymmetry)

DVJ: drop vertical jump; KMD: knee medial displacement; BMI: body mass index; ROM: range of motion; ADF_{KE}: ankle dorsiflexion with the knee extended; pVGRF: peak vertical ground reaction force; SLCMJ: single-leg countermovement jump; PHIR: passive hip internal rotation; FPPA: frontal plane projection angle; TJA: tuck jump assessment; BIL: bilateral ratio.

The baseline ZeroR classifier achieved an AUC of 0.5 ± 0, specificity of 100% and sensitivity of 0%. Table 3 shows the average AUC results for all resampling, ensemble and cost-sensitive learning methods separately for each decision base classifiers, nearly all of which have greater accuracy and sensitivity than the baseline model. As a result, a total of 3 algorithms built (using this subset of features) prediction models with AUC scores ≥0.7 (Table 4). Among these 3 algorithms, the UBAG with SMO as base classifier technique was the one that showed the highest F-score (0.380 ± 0.105) and hence, it was considered as the "best fit model". Therefore, the final screening model to prospectively classify male youth soccer players as having a high or low risk of suffering a LE-ST injury in the following 9 months of competitive season comprised 100 different SMO (rule-based) classifiers (an example of one of these SMO classifiers can be found in Figure 2, and the rest may be obtained upon request to the authors). In terms of practical applications, each classifier has a vote (yes [high risk of LE-ST injury] or no [lower risk of LE-ST injury]), and the final decision regarding whether or not a player might sustain an injury is determined by the combination of the votes of each individual classifier to each class (yes or no).

Table 3. AUC results (mean ± standard deviation) for the five base classifiers in isolation and after applying in them the resampling, ensemble (Classic, Boosting-based, Bagging-based, and Class-balanced ensembles), and cost-sensitive learning techniques selected.

Technique	Base classifiers									
	C4.5		ADTree		SMO		KNN		RF	
	AUC		AUC		AUC		AUC		AUC	
None	0.600	± 0.105	0.619	± 0.100	0.499	± 0.005	0.613	± 0.097	0.605	± 0.101
Resampling Techniques										
SMOTE	0.606	± 0.098	0.620	± 0.098	0.631	± 0.088	0.615	± 0.099	0.613	± 0.099
ROS	0.603	± 0.100	0.617	± 0.098	0.625	± 0.088	0.613	± 0.100	0.608	± 0.099
RUS	0.603	± 0.100	0.623	± 0.097	0.619	± 0.088	0.631	± 0.096	0.624	± 0.096
ENN	0.599	± 0.097	0.619	± 0.098	0.499	± 0.007	0.609	± 0.097	0.618	± 0.011
Classic Ensembles										
ADB1	0.636	± 0.091	0.614	± 0.098	0.610	± 0.076	0.575	± 0.106	-	-
M1	0.636	± 0.092	0.610	± 0.100	0.682	± 0.085	0.598	± 0.095	-	-
BAG	0.636	± 0.096	0.628	± 0.096	0.568	± 0.094	0.640	± 0.098	-	-
Boosting-based Ensembles										
SBO	0.614	± 0.097	0.611	± 0.100	0.671	± 0.091	0.609	± 0.095	-	-
RUSB	0.623	± 0.098	0.610	± 0.101	0.677	± 0.088	0.634	± 0.092	-	-
Bagging-based Ensembles										
OBAG	0.685	± 0.079	0.637	± 0.095	0.697	± 0.089	0.649	± 0.096	-	-
UBAG	0.653	± 0.089	0.631	± 0.096	**0.700**	**± 0.088**	0.667	± 0.091	-	-
SBAG	0.632	± 0.094	0.638	± 0.095	0.695	± 0.089	0.650	± 0.094	-	-
Cost-sensitive Classification										
MetaCost	0.577	± 0.099	0.623	± 0.103	0.500	± 0.011	0.604	± 0.097	-	-
CS-Classifier	0.597	± 0.101	0.618	± 0.098	0.539	± 0.066	0.621	± 0.096	-	-

Class-balanced Ensembles with a Cost-sensitive Classifier										
CS-OBAG	0.631	± 0.096	0.640	± 0.095	**0.704**	**± 0.085**	0.648	± 0.097	-	-
CS-UBAG	0.648	± 0.092	0.637	± 0.095	**0.703**	**± 0.084**	0.662	± 0.092	-	-
CS-SBAG	0.639	± 0.092	0.640	± 0.095	0.699	± 0.087	0.658	± 0.094	-	-

Highlighted in bold are the algorithms that built prediction models with AUC scores ≥0.7.

Table 4. Sub-set of algorithms that allowed building predictive models with AUC scores ≥0.7.

Technique	**Performance measures**							
	AUC		**TP rate (%)**		**TN rate (%)**		**F-score**	
UBAG [SMO]	**0.700**	**± 0.088**	**53.7**	**± 17.0**	**73.9**	**± 7.7**	**0.380**	**± 0.105**
CS-UBAG [SMO]	0.703	± 0.084	75.2	± 14.9	51.0	± 9.4	0.368	± 0.060
CS-OBAG [SMO]	0.704	± 0.085	72.8	± 15.2	55.1	± 9.3	0.379	± 0.066

Highlighted in bold is the algorithm with the highest F-score. AUC: area under the receiver operating characteristic curve; TP: true positive; TN: true negative.

Classifier 1 of the best fit prediction model for lower extremity non-contact soft tissue injuries (UBAG [SMO])

Equation → $f(x) = (w_1x_1 + ... + w_dx_d) + b = \langle w,x \rangle + b$

1. (-1.6022 * [normalized] BMI) +
2. (-2.2605 * [normalized] ROM-ADF$_{KE}$ dominant leg) +
3. (-1.4733 * [normalized] ROM-BIL-PHIR [No asymmetry]) +
4. (-0.6514 * [normalized] Landing BIL-pVGRF [SLCMJ] [No asymmetry]) +
5. (-1 * [normalized] BIL-FPPA [TJA] [No asymmetry]) +
6. (-0.075 * [normalized] KMD (dominant leg) [DVJ] [varus]) +
7. (-0.9167 * [normalized] KMD (dominant leg) [DVJ] [slight valgus]) +
8. (1.3451 * (normalized) KMD (dominant leg) [DVJ] [moderate valgus]) +
9. (-0.3534 * (normalized) KMD (dominant leg) [DVJ] [severe valgus]) +

5.3078 (b)

Classification:

- Negative score = Yes
- Positive score = No

Figure 2. Description of the first UBAG [SMO] classifier. BMI: body mass index; ROM: range of motion; ADFKE: ankle dorsiflexion with the knee extended; BIL: bilateral ratio; PHIR: passive hip internal rotation; pVGRF: peak vertical ground reaction force; SLCMJ: single-leg countermovement jump; FPPA: frontal plane projection angle; TJA: tuck jump assessment; KMD: knee medial displacement; DVJ: drop vertical jump.

For the model finally selected (UBAG with SMO as base classifier), an analysis of the average influence that each of its six variables has in the decision-making process regarding whether or not a player might suffer an injury was carried out by the SHAP approach and can be visualised in Figure 3. The variable that demonstrated the biggest impact was knee medial displacement (dominant leg) in the DVJ, followed by asymmetry in the peak vertical ground reaction force during landing in the SLCMJ, body mass index, asymmetry in the frontal plane projection angle assessed through the TJA, asymmetry in the passive hip internal rotation ROM, and ankle dorsiflexion with the knee extended (dominant leg) ROM. In Figure 4, the SHAP values for each feature value of an individual in the dataset are displayed.

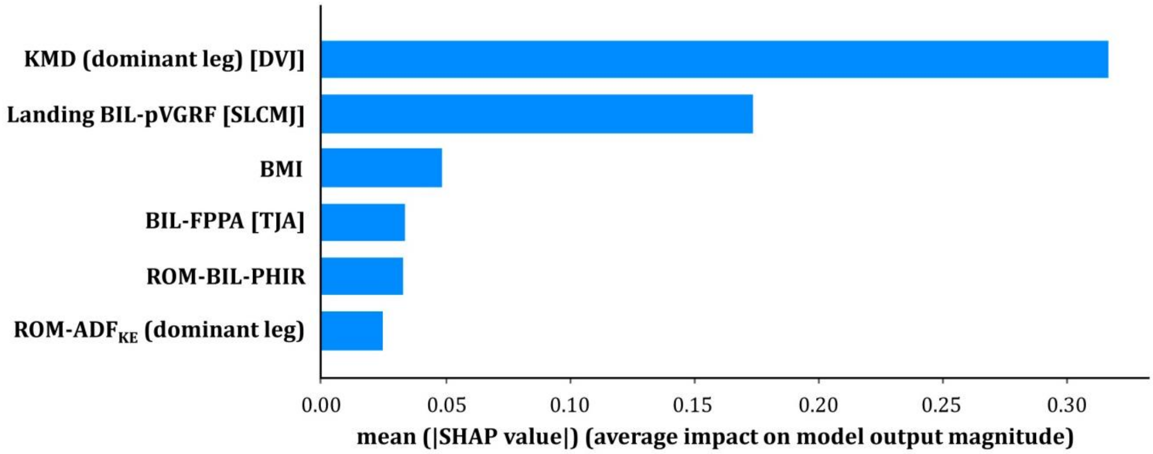

Figure 3. SHAP values for each feature. KMD: knee medial displacement; DVJ: drop vertical jump; BIL: bilateral ratio; pVGRF: peak vertical ground reaction force; SLCMJ: single-leg countermovement jump; BMI: body mass index; FPPA: frontal plane projection angle; TJA: tuck jump assessment; ROM: range of motion; PHIR: passive hip internal rotation; ADFKE: ankle dorsiflexion with the knee extended.

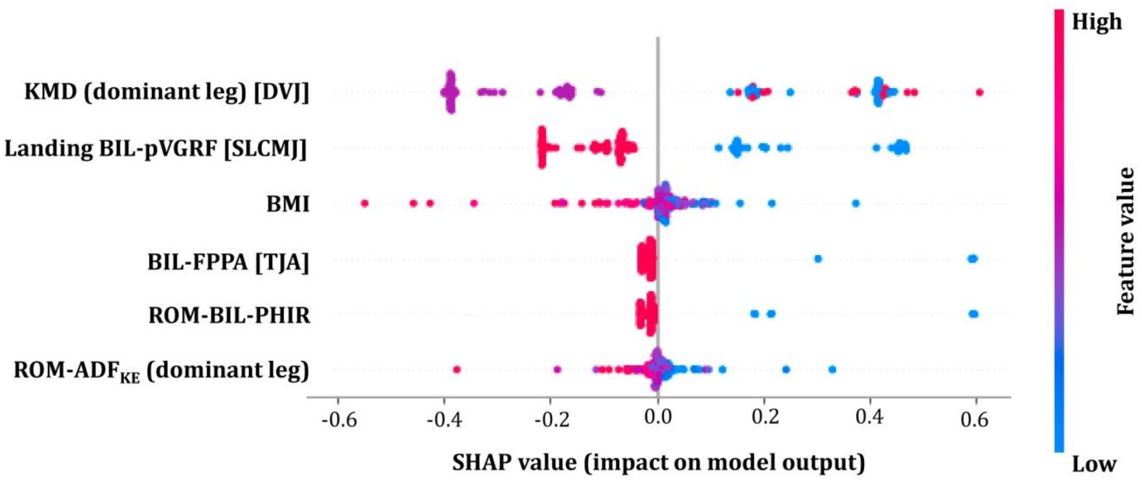

Figure 4. SHAP summary plot. The features in the model are listed from the most (top) to least (bottom) important by their global impact on the model. Dots representing the SHAP values for each feature value of an individual in the dataset are plotted horizontally next to the feature. Overlapping points are jittered in y-axis direction, so a sense of the distribution of the Shapley values per variable is achieved. The higher the absolute value (either positive or negative), the higher the importance in the

classification decision-making process. Positive SHAP values represent a higher probability of a negative prediction (i.e., No injured). Each dot is colored by the value (i.e., measured value) of the feature for an individual, where blue represents the lower values (e.g., lower BMI score) and red the higher values (e.g., higher BMI scores).

4 Discussion

The aim of this study was twofold: a) to build models using machine learning techniques on data from an extensive screening battery to prospectively predict LE-ST injuries in non-elite male youth soccer players, and b) to compare their performance scores (i.e., accuracy) to select the best fit prediction model. In this sense, the present study has built a screening model (AUC = 0.700) based on six pre-season field-based measures to predict LE-ST injuries in male youth soccer players. In particular, the model developed successfully identifies one out of every two (TP rate = 53.7%) and three out of every four (TN rate = 73.9%) male youth soccer players at high or low risk of suffering a LE-ST injury throughout the in-season phase, respectively.

The ability of the derived model in the current study to predict LE-ST injuries is similar to the model developed by Oliver et al. [14] (AUC = 0.663, TP rate = 55.6%, TN rate = 74.2%) but lower than the model reported by Rommers et al. [21] (AUC = 0.850, TP rate = 85%), albeit both using elite-level male youth soccer players. Three different arguments may explain the higher performance scores reported by Rommers et al.'s [21] model compared to those shown in the current prediction models and that built by Oliver et al. [14]:

The first argument that may be used to explain these differences in the models' performance is the larger number of players that were enrolled in the study conducted by Rommers et al. [21] (*n* = 734) in comparison with Oliver et al.'s [14] study (*n* = 355) and the current research (*n* = 260). In studies dealing with class imbalance problems, such as the LE-ST injury phenomenon, in which the number of injured players (minority class) prospectively reported is always much lower than the non-injured participants (majority class) [19,51], large sample sizes may be required to ensure having enough instances in the minority class to avoid them being considered as noise by the learning algorithms during the process of building models. In this sense, Japkowicz & Stephen [52] demonstrated that the error rate caused by imbalanced class distribution decreases when the number of examples of the minority class is representative. While Rommers et al. [21] identified 368 injured players throughout the follow up, Oliver et al. [14] and the current study used 99 and 45 injuries respectively to develop the prediction models. Therefore, in the model built by Rommers et al. [21], patterns that were defined by injury players could have been better learned and this may have positively impacted on its predictive ability.

The second argument is linked to the fact that the imbalance ratios (IR = ∑injured players / ∑non-injured players) of the dichotomic class variable (injury yes or no) in Oliver et al.'s [14] study (IR = 0.39) and the current study (IR = 0.21) were much higher than the one observed in Rommers et al.'s [21] study (IR = 1.00). In fact, the data set used by Rommers et al. [21] to build their injury prediction model did not show an imbalanced distribution in the class variable as the number of injured (*n* = 368) and non-injured (*n* = 366) players was almost the same. Class distribution (i.e., the proportion of instances [e.g., soccer players] belonging to each class [injured vs non-injured] in a data-set) plays a key role in classification problems. Highly imbalance data sets usually tend to suffer from class overlapping and/or small disjuncts, which difficult classifier learning [51]. Thus, although Oliver et al. [14] and the current study have used learning algorithms specially designed to deal with class imbalance problems and acceptable predictive accuracy results were reported, the lower IR in the study of Rommers et al. [21] may have allowed lower misclassification rates and hence, better accuracy scores. In this sense, the different weekly exposure (in terms of frequency and physical demands) to the soccer play that could have occurred between our sample of amateur youth soccer players and the elite ones used by Rommers et al [21] and Oliver et al. [14] might be one of the main reasons for the lower injury rates and consequently the higher imbalance ratio found in the current research. The participants of our study routinely completed a total of two (U11-12 players) and three (U13-14, U15-16, and U17-19 players) 90-min training sessions per week on non-consecutive days and played one competitive match usually at the weekend. In addition, in all age groups, the competitive season was divided into three blocks of 9-12 weeks separated by a 2-3-week break (coinciding with Christmas and Easter festivities). On the contrary, it is plausible that the elite youth soccer players (mainly those belonging to the more advanced age groups) who took part in both Oliver et al.'s [14] and Rommers et al.'s [21] studies could have shown larger (i.e., number of training session per week) and higher physically demanding weekly exposures to the game of soccer than our participants. This higher frequency and intensity in the exposure to soccer that usually elite adolescent (>14 years old) players have in comparison with their counterpart non-elite players might be attributed to the early sport specialisation process that usually is observed in the youth academies of professional soccer clubs. Furthermore, it is also possible that the participants in Oliver et al.'s [14] and Rommers et al.'s [21] studies may have had shorter Christmas and Easter breaks (in case they had any of them) than our amateur youth soccer players. Therefore, the larger and higher physically demanding weekly exposure alongside the shorter resting periods that may have had the elite youth soccer players that took part in these two studies may have led them to a progressive and chronic accumulation of fatigue that could have dramatically increased their risk of injury. However, as neither Oliver et al. [14] nor Rommers et al. [21] reported the weekly exposure to the game of soccer in their participants, this hypothesis should be considered with a degree of caution.

Finally, the last aspect that might have also played a key role in the higher predictive ability observed in the model published by Rommers et al. [21] is the less exigent resampling method applied to determine its ability to predict injuries. In particular, Rommers et al. [21] used a hold out with 20% of the same as test data to assess the predictive ability of its model whereas Oliver et al. [14] employed a five-fold cross validation technique and the present study repeated 100 times this five-fold cross validation procedure in an attempt to achieve a more accurate estimation of the models' performance. It has been suggested that the k-fold cross validation is a more powerful preventive technique against model performance overfitting than the hold out because the validation metrics calculated for each fold are combined to give an overall estimate of the model's performance, reducing the risk of accidentally obtaining a really optimistic test data [53]. Unlike the current study, neither Rommers et al. [21] nor Oliver et al. [14] uploaded their respective data sets into a public repository. Consequently, we were not able to apply the resampling technique used in the current study to assess the prediction ability of their models in order to inform whether (or not) their performance scores could have suffered from overfitting (and specify to some extent) due to a less exigent validation technique.

Another main finding of the current study is that of the 135 potential risk factors obtained from the several questionnaires and field-based tests carried out during the pre-season testing session conducted in each soccer team, only six (Table 2) were finally selected as the most important features related to LE-ST injuries. This subset comprised of an anthropometric parameter (BMI), three neuromuscular measures (KMD in the dominant leg [DVJ], landing BIL-pVGRF [SLCMJ] and BIL-FPPA [TJA]) and two joint ROMs (ROM-BIL-PHIR and ROM-ADF_{KE} in the dominant leg) allowed us to build a model to predict LE-ST injuries in male youth soccer players. Therefore, one of the main advantages of the model presented in this study is that it only needs five to ten minutes to run the screen in a single player, unlike Rommers et al.'s [21] model where 20 measures recorded from a questionnaire and five different field-based tests are required, which can take longer than 45 min to collect all data in a single athlete. The six measures selected have been consistently proposed as primary injury risk factors for LE-ST injuries in several prospective and biomechanical studies conducted in paediatric athlete population [14,54,55]. As it is shown in Figure 3, a higher knee medial displacement (i.e., dynamic knee valgus) of the dominant leg in DVJ (SHAP score = 0.32) and the presence of asymmetries in pVGRF at landing from SLCMJ (SHAP score = 0.17) were identified as the two most important predictors for LE-ST injury. A higher body mass index (SHAP score = 0.05), bilateral differences ≥ 10% in FPPA measured through the TJA manoeuvre (SHAP score = 0.03) and ≥ 8º in PHIR ROM (SHAP score = 0.03), and lower ADF_{KE} ROM of the dominant leg (SHAP score = 0.02) had a smaller effect on the prediction model. It is beyond the scope of this study to describe into detail the potential mechanisms that justify the reasons why each of these six measures themselves might increase the vulnerability to LE-ST injury in this cohort of soccer players. However, the proposed mechanisms might include altered frontal (i.e., the adoption of an excessive dynamic valgus motion at the knee [high KMD and FPPA scores]), sagittal (ankle ROM) and transverse (hip internal rotation ROM) planes during the execution of high intensity weight-bearing dynamic tasks (e.g., landing from a jump, side-stepping, pressing and tackling) that may produce increased loading of the knee and ankle

[54,56]. Likewise, it has been suggested that increased BMI scores may imply changes in moments of inertia, forces and deformations experienced by various soft tissues during high intensity movements (e.g., high speed running, change of direction) [4], which may be associated with injury risk, particularly muscle strains [55]. Asymmetries in pVGRF at landing from SLCMJ have been also identified by previous studies as a primary injury risk factor in male youth soccer players [14] and it is deemed to place additional stress on the weaker leg predisposing it to increased injury risk. Importantly, these six measures are considered modifiable risk factors and hence, some strategies can be implemented to optimise these factors in each player to lower the probability of suffering a LE-ST injury. In this regard, previous studies have demonstrated that the regular application of short (not more than 20-25 min) bouts of multi-component exercises during training sessions can significantly improve, among other aspects, neuromuscular control and performance and help to control body weight in team sport athletes (including young soccer players) [42,57]. Therefore, these multi-component programs may be powerful tools to be used by practitioners as preventive measures in those soccer players categorised at high risk of LE-ST injury.

Finally, it should be highlighted that simulations ran in our laboratory showed that giving the four basic algorithms used in this study (C4.5, ADTree, SMO and KNN) the opportunity to select by themselves (according to their own criteria) the most relevant variables did not improve the predictive performance of the models but increased its complexity. Furthermore, simulations were also run with other attribute evaluators (such as InfoGain and Correlations) to select relevant features and none of them improved the performance scores presented in this study.

4.1 Limitations

This study has also some limitations that should be acknowledged. Even though all the variables collected during the screening session are considered as risk factors for LE-ST injuries, there are additional measures from various questionnaires and field-based tests that were not assessed in this research (due to time restrictions) and that may have enhanced the ability to predict LE-ST injuries in this cohort of young athletes (e.g., trunk stability measures, relative leg stiffness, and change of direction kinematics). Likewise, the complex interaction of growth, maturity timing and tempo across players of varying age and maturity along with the fact that a non-single type of injury (e.g., hamstring strains, ACL tears) was analysed may have reduced the ability of the feature selection algorithm applied to the data set to reduce its dimensionality (through removing redundant and not relevant measures), and thus could have penalised the performance of the model. Future studies should assess whether (or not) the use of more homogeneous samples, in terms of maturity status, and focusing the attention on single types of injury may increase the predictive ability of the screening models. Another limitation of the current study is that only the first occurring injury of every player was considered in the analysis. Consequently, because players can sustain multiple injuries over one season, the analysis does not reflect the complete picture. Furthermore, players were only tested at the end of the pre-season with subsequent injuries monitored over the entire season. Anthropometric, physical fitness, neuromuscular capability and biomechanical measures change over the course of the season due to training and natural development [21,55], which may have negatively impacted on the models ability to predict injuries. Therefore, future studies should conduct screening session every few months in order to obtain accurate screening data that is closer to the time of injury, mitigating the effects of training, growth and maturation.

5 Conclusions

Due to the application of machine learning techniques, the current study has developed a screening model based on six field-based measures that showed moderate validity (AUC score = 0.700, TPrate = 53.7% and TNrate = 73.9% determined through the exigent repeated cross-validation resampling technique) for identifying youth soccer players at risk of LE-ST injury. Furthermore, and thanks to the SHAP approach, it is possible to determine the influence of each risk factor selected (i.e., KMD [dominant leg] in the DVJ, landing BIL-pVGRF [SLCMJ], BMI, BIL-FPPA [TJA], ROM-BIL-PHIR and ROM-ADF$_{KE}$ [dominant leg]) in the prediction model (injury yes vs. injury no). Given that these measures require little equipment to be obtained and can be employed quickly (approximately 5-10 min) and easily by trained staff in a single player, the model developed in this study might be included as an essential component of the injury management strategy in youth soccer.

Acknowledgements

The authors would like to thank the participating players and coaches for their collaboration in this study, and the assessment team for their support in the data collection.

Authors' Contributions

FJR-P, MDSC, PSdB, and FA conceived and designed the research; FJR-P, AC, FS, and FA obtained the data; FJR-P, JMP-C, JAG, and FA analysed and interpreted the data; FJR-P, PSdB, and FA led the drafting of the manuscript; JMP-C, JAG, MDSC, AC, and FS revised the manuscript critically for important intellectual content. All authors have read and approved the final version of the manuscript, and agree with the order of presentation of the authors.

Funding

Francisco Javier Robles-Palazón was initially supported by the Program of Human Resources Formation for Science and Technology (20326/FPI/2017) from the Seneca Foundation-Agency for Science and Technology in the Region of Murcia (Spain) and subsequently by a Margarita Salas postdoctoral fellowship (UMU/R-1500/2021) given by the Spanish Ministry of Universities and funded by the European Union–NextGenerationEU. Francisco Ayala was supported by a Ramón y Cajal postdoctoral fellowship given by the Spanish Ministry of Science and Innovation (RYC2019-028383-I). This study is part of the projects entitled "Estudio del riesgo de lesión en jóvenes deportistas a través de redes de inteligencia artificial" (DEP2017-88775-P) and "El fútbol femenino importa: identificación del riesgo de lesión a través de la inteligencia artificial" (PID2020-115886RB-I00), funded by the Spanish Ministry of Science and Innovation, the State Research Agency (AEI) and the European Regional Development Fund (ERDF). The funders had no role in study design, data analysis, interpretation, or the decision to submit the work for publication.

Availability of Data and Materials

The scheme of the algorithm finally selected is displayed in online supplementary file 9 and the model is publicly available on https://data.mendeley.com/datasets/2mw6w556yg/1.

Declarations

Ethics Approval and Consent to Participate

The experimental procedures used in this study were in accordance with the Declaration of Helsinki and were approved by the Ethics and Scientific Committee of the University of Murcia, Spain (ID: 1551/2017). All participants provided written informed consent prior to the study.

Competing Interests

Francisco Javier Robles-Palazón, José M. Puerta-Callejón, José A. Gámez, Mark De Ste Croix, Antonio Cejudo, Fernando Santonja, Pilar Sainz de Baranda, and Francisco Ayala declare that they have no competing interests.

References

[1] Robles-Palazón FJ, López-Valenciano A, De Ste Croix MB, Oliver JL, García-Gómez A, Sainz de Baranda P, et al. Epidemiology of injuries in male and female youth football players: a systematic review and meta-analysis. J Sport Health Sci 2022;11:681–695.

[2] Bult HJ, Barendrecht M, Tak IJR. Injury risk and injury burden are related to age group and peak height velocity among talented male youth soccer players. Orthop J Sports Med 2018;6:1–10. https://doi.org/10.1177/2325967118811042.

[3] van der Sluis A, Elferink-Gemser MT, Coelho-e-Silva MJ, Nijboer JA, Brink MS, Visscher C. Sport injuries aligned to peak height velocity in talented pubertal soccer players. Int J Sports Med 2014;35:351–5. https://doi.org/10.1055/s-0033-1349874.

[4] Hawkins D, Metheny J. Overuse injuries in youth sports: biomechanical considerations. Med

Sci Sport Exerc 2001;33:1701–7.

[5] Philippaerts RM, Vaeyens R, Janssens M, Van Renterghem B, Matthys D, Craen R, et al. The relationship between peak height velocity and physical performance in youth soccer players. J Sports Sci 2006;24:221–30. https://doi.org/10.1080/02640410500189371.

[6] Wik EH, Lolli L, Chamari K, Materne O, Di Salvo V, Gregson W, et al. Injury patterns differ with age in male youth football: a four-season prospective study of 1111 time-loss injuries in an elite national academy. Br J Sports Med 2021;55:794–800. https://doi.org/10.1136/bjsports-2020-103430.

[7] Verhagen E, Bolling C, Finch CF. Caution this drug may cause serious harm! why we must report adverse effects of physical activity promotion. Br J Sports Med 2015;49:1–2. https://doi.org/10.1136/bjsports-2014-093604.

[8] Maffulli N, Longo UG, Gougoulias N, Loppini M, Denaro V. Long-term health outcomes of youth sports injuries. Br J Sports Med 2010;44:21–5. https://doi.org/10.1136/bjsm.2009.069526.

[9] Rössler R, Junge A, Bizzini M, Verhagen E, Chomiak J, aus der Fünten K, et al. A multinational cluster randomised controlled trial to assess the efficacy of “11 + Kids”: a warm-up programme to prevent injuries in children’s football. Sports Med 2018;48:1493–504. https://doi.org/10.1007/s40279-017-0834-8.

[10] Bahr R. Why screening tests to predict injury do not work—and probably never will...: a critical review. Br J Sports Med 2016;50:776–80. https://doi.org/10.1136/bjsports-2016-096256.

[11] Ayala F, López-Valenciano A, Gámez Martín JA, De Ste Croix M, Vera-Garcia FJ, García-Vaquero M del P, et al. A preventive model for hamstring injuries in professional soccer: learning algorithms. Int J Sports Med 2019;40:344–53.

[12] Ruiz-Pérez I, López-Valenciano A, Hernández-Sánchez S, Puerta-Callejón JM, De Ste Croix M, Sainz de Baranda P, et al. A field-based approach to determine soft tissue injury risk in elite futsal using novel machine learning techniques. Front Psychol 2021;12:1–15.

https://doi.org/10.3389/fpsyg.2021.610210.

[13] López-Valenciano A, Ayala F, Puerta JM, De Ste Croix M, Vera-García F, Hernández-Sánchez S, et al. A preventive model for muscle injuries: a novel approach based on learning algorithms. Med Sci Sports Exerc 2018;50:915–27. https://doi.org/10.1016/j.physbeh.2017.03.040.

[14] Oliver JL, Ayala F, De Ste Croix MBA, Lloyd RS, Myer GD, Read PJ. Using machine learning to improve our understanding of injury risk and prediction in elite male youth football players. J Sci Med Sport 2020;23:1044–8. https://doi.org/10.1016/j.jsams.2020.04.021.

[15] Robles-Palazón FJ, Ruiz-Pérez I, Aparicio-Sarmiento A, Cejudo A, Ayala F, Sainz de Baranda P. Incidence, burden, and pattern of injuries in Spanish male youth soccer players: A prospective cohort study. Phys Ther Sport 2022;56:48–59. https://doi.org/10.1016/j.ptsp.2022.06.005.

[16] De Ridder R, Witvrouw E, Dolphens M, Roosen P, Van Ginckel A. Hip strength as an intrinsic risk factor for lateral ankle sprains in youth soccer players. Am J Sports Med 2017;45:410–6.

[17] Ko J, Rosen AB, Brown CN. Functional performance tests identify lateral ankle sprain risk: a prospective pilot study in adolescent soccer players. Scand J Med Sci Sport 2018;28:2611–6. https://doi.org/10.1111/sms.13279.

[18] Padua DA, DiStefano LJ, Beutler AI, de la Motte SJ, DiStefano MJ, Marshall SW. The landing error scoring system as a screening tool for an anterior cruciate ligament injury-prevention program in elite-youth soccer athletes. J Athl Train 2015;50:589–95. https://doi.org/10.4085/1062-6050-50.1.10.

[19] López V, Fernández A, García S, Palade V, Herrera F. An insight into classification with imbalanced data: empirical results and current trends on using data intrinsic characteristics. Inf Sci (Ny) 2013;250:113–41. https://doi.org/https://doi.org/10.1016/j.ins.2013.07.007.

[20] Bittencourt NFN, Meeuwisse WH, Mendonça LD, Nettel-Aguirre A, Ocarino JM, Fonseca ST. Complex systems approach for sports injuries: moving from risk factor identification to injury pattern recognition - narrative review and new concept. Br J Sports Med 2016;50:1309–14.

https://doi.org/10.1136/bjsports-2015-095850.

[21] Rommers N, Rössler R, Verhagen E, Vandecasteele F, Verstockt S, Vaeyens R, et al. A machine learning approach to assess injury risk in elite youth football players. Med Sci Sport Exerc 2020;52:1745–51. https://doi.org/10.1249/mss.0000000000002305.

[22] Vaeyens R, Malina RM, Janssens M, Van Renterghem B, Bourgois J, Vrijens J, et al. A multidisciplinary selection model for youth soccer: the Ghent Youth Soccer Project. Br J Sports Med 2006;40:928–34. https://doi.org/10.1136/bjsm.2006.029652.

[23] Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. Ann Intern Med 2015;162:55–63. https://doi.org/10.7326/M14-0697.

[24] Cejudo A, Sainz de Baranda P, Ayala F, De Ste Croix M, Santonja-Medina F. Assessment of the range of movement of the lower limb in sport: advantages of the Rom-Sport I battery. Int J Environ Res Public Health 2020;17:7606. https://doi.org/10.3390/ijerph17207606.

[25] Shaffer SW, Teyhen DS, Lorenson CL, Warren RL, Koreerat CM, Straseske CA, et al. Y-balance test: a reliability study involving multiple raters. Mil Med 2013;178:1264–70. https://doi.org/10.7205/MILMED-D-13-00222.

[26] Romero-Franco N, Jiménez-Reyes P, Castaño-Zambudio A, Capelo-Ramírez F, Rodríguez-Juan JJ, González-Hernández J, et al. Sprint performance and mechanical outputs computed with an iPhone app: comparison with existing reference methods. Eur J Sport Sci 2017;17:386–92. https://doi.org/10.1080/17461391.2016.1249031.

[27] Read PJ, Oliver JL, De Ste Croix MBA, Myer GD, Lloyd RS. Landing kinematics in elite male youth soccer players of different chronologic ages and stages of maturation. J Athl Train 2018;53:372–8. https://doi.org/10.4085/1062-6050-493-16.

[28] Ortiz A, Rosario-Canales M, Rodríguez A, Seda A, Figueroa C, Venegas-Ríos H. Reliability and concurrent validity between two-dimensional and three-dimensional evaluations of knee valgus during drop jumps. Open Access J Sports Med 2016;7:65–73. https://doi.org/10.2147/oajsm.s100242.

[29] Robles-Palazón FJ, Ruiz-Pérez I, Oliver JL, Ayala F, Sainz de Baranda P. Reliability, validity, and maturation-related differences of frontal and sagittal plane landing kinematic measures during drop jump and tuck jump screening tests in male youth soccer players. Phys Ther Sport 2021;50:206–16. https://doi.org/10.1016/j.ptsp.2021.05.009.

[30] Myer GD, Ford KR, Hewett TE. New method to identify athletes at high risk of ACL injury using clinic-based measurements and freeware computer analysis. Br J Sports Med 2011;45:238–44. https://doi.org/10.1136/bjsm.2010.072843.

[31] Mirwald RL, Baxter-Jones ADG, Bailey DA, Beunen GP. An assessment of maturity from anthropometric measurements. Med Sci Sports Exerc 2002;34:689–94. https://doi.org/10.1249/00005768-200204000-00020.

[32] Buela-Casal G, Guillén-Riquelme A, Seisdedos Cubero N. Cuestionario de ansiedad estado-rasgo: adaptación española. Madrid: TEA Ediciones; 2011.

[33] Andrade E, Arce C, Armental J, Rodríguez M, de Francisco C. Indicadores del estado de ánimo en deportistas adolescentes según el modelo multidimensional del POMS. Psicothema 2008;20:630–5.

[34] Gimeno F, Buceta JM, Pérez-Llanta M. El cuestionario «Características psicológicas relacionadas con el rendimiento deportivo»(CPRD): características psicométricas. Análise Psicológica 2001;19:93–113.

[35] Taylor KL, Sheppard JM, Lee H, Plummer N. Negative effect of static stretching restored when combined with a sport specific warm-up component. J Sci Med Sport 2009;12:657–61. https://doi.org/10.1016/j.jsams.2008.04.004.

[36] Myer GD, Ford KR, Hewett TE. Tuck jump assessment for reducing anterior cruciate ligament injury risk. Athl Ther Today 2008;13:39–44.

[37] Thomas K, French D, Hayes PR. The effect of two plyometric training techniques on muscular power and agility in youth soccer players. J Strength Cond Res 2009;23:332–5.

[38] Street G, McMillan S, Board W, Rasmussen M, Heneghan JM. Sources of error in determining countermovement jump height with the impulse method. J Appl Biomech 2001;17:43–54.

https://doi.org/10.1123/jab.17.1.43.

[39] Grindem H, Logerstedt D, Eitzen I, Moksnes H, Axe MJ, Snyder-Mackler L, et al. Single-legged Hop tests as predictors of self-reported knee function in nonoperatively treated individuals with anterior cruciate ligament injury. Am J Sports Med 2011;39:2347–54. https://doi.org/10.1177/0363546511417085.

[40] Runacres A, Bezodis NE, Mackintosh KA, McNarry MA. The reliability of force-velocity-power profiling during over-ground sprinting in children and adolescents. J Sports Sci 2019;37:2131–7. https://doi.org/10.1080/02640414.2019.1622316.

[41] Robles-Palazón FJ, Ayala F, Cejudo A, De Ste Croix M, Sainz de Baranda P, Santonja F. Effects of age and maturation on lower extremity range of motion in male youth soccer players. J Strength Cond Res 2022;36:1417–25. https://doi.org/10.1519/JSC.0000000000003642.

[42] Faude O, Rössler R, Petushek EJ, Roth R, Zahner L, Donath L. Neuromuscular adaptations to multimodal injury prevention programs in youth sports: a systematic review with meta-analysis of randomized controlled trials. Front Physiol 2017;8:791. https://doi.org/10.3389/fphys.2017.00791.

[43] Fuller CW, Ekstrand J, Junge A, Andersen TE, Bahr R, Dvorak J, et al. Consensus statement on injury definitions and data collection procedures in studies of football (soccer) injuries. Br J Sports Med 2006;40:193–201. https://doi.org/10.1136/bjsm.2005.025270.

[44] Paulsen A, Overgaard S, Lauritsen JM. Quality of data entry using single entry, double entry and automated forms processing-an example based on a study of patient-reported outcomes. PLoS One 2012;7:1–6. https://doi.org/10.1371/journal.pone.0035087.

[45] Altman DG, Bland JM. Diagnostic tests 3: receiver operating characteristic plots. BMJ 1994;309:188. https://doi.org/10.1136/bmj.309.6947.102.

[46] Zou Q, Xie S, Lin Z, Wu M, Ju Y. Finding the best classification threshold in imbalanced classification. Big Data Res 2016;5:2–8. https://doi.org/https://doi.org/10.1016/j.bdr.2015.12.001.

[47] Elkarami B, Alkhateeb A, Rueda L. Cost-sensitive classification on class-balanced ensembles

for imbalanced non-coding RNA data. 2016 IEEE EMBS Int. Student Conf. ISC, IEEE; 2016, p. 1–4. https://doi.org/10.1109/EMBSISC.2016.7508607.

[48] Robertson PS. Man & machine: Adaptive tools for the contemporary performance analyst. J Sports Sci 2020;38:2118–26. https://doi.org/10.1080/02640414.2020.1774143.

[49] Hall MA, Smith LA. Practical feature subset selection for machine learning. Comput. Sci. '98 Proc. 21st Australas. Comput. Sci. Conf. ACSC'98, 1998, p. 181–91.

[50] Lundberg SM, Erion GG, Lee SI. Consistent individualized feature attribution for tree ensembles. ArXiv 2018:1802.03888.

[51] Galar M, Fernandez A, Barrenechea E, Bustince H, Herrera F. A review on ensembles for the class imbalance problem: bagging-, boosting-, and hybrid-based approaches. IEEE Trans Syst Man, Cybern Part C (Applications Rev 2012;42:463–84. https://doi.org/10.1109/TSMCC.2011.2161285.

[52] Japkowicz N, Stephen S. The class imbalance problem: a systematic study. Intell Data Anal 2002;6:429–49. https://doi.org/10.3233/IDA-2002-6504.

[53] Yadav S, Shukla S. Analysis of k-fold cross-validation over hold-out validation on colossal datasets for quality classification. 2016 IEEE 6th Int. Conf. Adv. Comput., IEEE; 2016, p. 78–83. https://doi.org/10.1109/IACC.2016.25.

[54] Hewett TE, Myer GD, Ford KR, Heidt RS, Colosimo AJ, McLean SG, et al. Biomechanical measures of neuromuscular control and valgus loading of the knee predict anterior cruciate ligament injury risk in female athletes: a prospective study. Am J Sports Med 2005;33:492–501. https://doi.org/10.1177/0363546504269591.

[55] Kemper GLJ, van der Sluis A, Brink MS, Visscher C, Frencken WGP, Elferink-Gemser MT. Anthropometric injury risk factors in elite-standard youth soccer. Int J Sports Med 2015;36:1112–7. https://doi.org/10.1055/s-0035-1555778.

[56] Koga H, Nakamae A, Shima Y, Bahr R, Krosshaug T. Hip and ankle kinematics in noncontact anterior cruciate ligament injury situations: video analysis using model-based image matching. Am J Sports Med 2018;46:333–40. https://doi.org/10.1177/0363546517732750.

[57] McHugh MP. Oversized young athletes: a weighty concern. Br J Sports Med 2010;44:45–9. https://doi.org/10.1136/bjsm.2009.068106.