Received: 11 February 2022 | Revised: 17 January 2023 | Accepted: 22 January 2023

DOI: 10.1111/sms.14322

ORIGINAL ARTICLE

WILEY

Predictive modeling of lower extremity injury risk in male elite youth soccer players using least absolute shrinkage and selection operator regression

Mathias Kolodziej[1,2] | Andreas Groll[3] | Kevin Nolte[2] | Steffen Willwacher[4] | Tobias Alt[5] | Marcus Schmidt[2] | Thomas Jaitner[2]

[1]Department of Strength and Conditioning and Performance, Borussia Dortmund, Dortmund, Germany

[2]Institute for Sports and Sports Science, TU Dortmund University, Dortmund, Germany

[3]Department of Statistics, TU Dortmund University, Dortmund, Germany

[4]Department of Mechanical and Process Engineering, Offenburg University of Applied Sciences, Offenburg, Germany

[5]Department of Biomechanics, Performance Analysis and Strength and Conditioning, Olympic Training and Testing Centre Westphalia, Dortmund, Germany

Correspondence
Mathias Kolodziej, Department of Strength and Conditioning and Performance, Borussia Dortmund, Dortmund, Germany.
Email: mathias.kolodziej@bvb.de

Purpose: To (1) identify neuromuscular and biomechanical injury risk factors in elite youth soccer players and (2) assess the predictive ability of a machine learning approach.

Material and Methods: Fifty-six elite male youth soccer players (age: 17.2 ± 1.1 years; height: 179 ± 8 cm; mass: 70.4 ± 9.2 kg) performed a 3D motion analysis, postural control testing, and strength testing. Non-contact lower extremities injuries were documented throughout 10 months. A least absolute shrinkage and selection operator (LASSO) regression model was used to identify the most important injury predictors. Predictive performance of the LASSO model was determined in a leave-one-out (LOO) prediction competition.

Results: Twenty-three non-contact injuries were registered. The LASSO model identified concentric knee extensor peak torque, hip transversal plane moment in the single-leg drop landing task and center of pressure sway in the single-leg stance test as the three most important predictors for injury in that order. The LASSO model was able to predict injury outcomes with a likelihood of 58% and an area under the ROC curve of 0.63 (sensitivity = 35%; specificity = 79%).

Conclusion: The three most important variables for predicting the injury outcome suggest the importance of neuromuscular and biomechanical performance measures in elite youth soccer. These preliminary results may have practical implications for future directions in injury risk screening and planning, as well as for the development of customized training programs to counteract intrinsic injury risk factors. However, the poor predictive performance of the final model confirms the challenge of predicting sports injuries, and the model must therefore be evaluated in larger samples.

KEYWORDS

adolescent, elite, injury prediction, laboratory-based injury risk screening, soccer

1 | INTRODUCTION

Male elite youth soccer players boast a high injury risk, especially to the lower extremities, which is often related to early specialization and high training loads.[1,2] Additionally, injuries and the resulting reduced player availability negatively impact team and individual performances.[3] Thus, the challenge of predicting injury risk has received growing attention in recent years, not only in sport science but also in the sports industry,[4] mainly since the identification of injury risk factors and the understanding of their interplay represent an important foundation for effective injury prevention.[5]

The cause-effect paradigm has been widely applied in injury prevention research, with such approaches focusing on finding the association between a single risk factor and injury outcomes (univariate approach). Research using this methodological approach is essential in terms of understanding why an injury occurs.[6] Hence, several modifiable and non-modifiable risk factors have been hypothesized for lower extremity injuries using a multitude of statistical methods.[7] However, although an injury may occur due to a single risk factor, this determination displays only a small piece of the puzzle.[8] In reality, the occurrence of an injury represents a complex systemic reaction and tends to be more likely the result of a non-linear interaction between multiple risk factors (multivariable approach).[5,6,8,9] Additionally, predictive modeling of lower extremity injury risk should focus on forecasting the occurrence of an injury and making predictions from known values to unknown outcomes.[8]

In recent years, the application of more advanced contemporary statistical approaches (e.g., supervised learning algorithms) derived from advanced artificial intelligence has emerged in sports injury prediction research to tackle this challenging multi-faceted task.[4,10] Traditional logistic regression is often used to analyze the ability of multiple risk factors, which are determined through injury risk screening batteries, to predict injuries.[5,6,11] However, if a large number of possible predictors is available, standard logistic regression estimates often become unstable or even infeasible ($p > n$ case). Moreover, and especially if multiple interaction effects of the predictors are also included in the statistical model, multicollinearity issues arise. In these situations, regularization techniques and methods for variable selection become relevant. Besides other methods stemming from the machine learning community, such as boosting[12] or random forests,[13] the so-called least absolute shrinkage and selection operator (LASSO[14,15]) has proven to be particularly beneficial.[4] An alternative approach for the selection of sensitive parameter subsets in the context of the biomechanical system identification has been proposed by Ramadan et al.[16]

Recent empirical evidence has shown that contemporary statistical approaches, including machine learning, can provide promising results in the prediction of injuries in multiple sports using a variety of predictor variables.[11,17,18] Two of the most promising, but also contrasting, sets of results to predict injury risk were obtained in elite youth soccer players.[11,17]

Rommers et al[17] and Oliver et al[11] both prospectively investigated a large sample of male youth elite soccer players (Rommers et al[17] examined U10–U15 age groups, whereas Olivier et al[11] examined U11–U18 age groups) through a preseason test battery of anthropometric, motor coordination, and physical measures.[11,17] Rommers et al[17] identified a higher predicted age at peak height velocity, higher body height, longer leg length, lower fat percentage, and average performance on the standing broad jump as the five most important predictors for injury, using an extreme gradient boosting algorithm (XGBoost) with an accuracy of 85%.

Based on their machine learning model, which used the best performing decision tree, Oliver et al[11] concluded that asymmetry in the single-leg countermovement jump, asymmetry in the 75% Hop, asymmetry in the Y-Balance Test, the knee valgus angle assessed through the tuck jump test, and body size contribute to injury risk. In contrast to Rommers et al,[17] their machine learning model had poor overall accuracy in detecting injury.

Despite the different outcomes, both studies used field-based and low-cost screening tests—in addition to anthropometric measures—whose measurement properties and relationships to injury are currently limited, conflicting, or unknown.[19] Additionally, and to the best of the authors' knowledge, previous machine learning approaches have been limited in their inclusion of measurements of muscle strength, postural stability, and biomechanics of the lower extremities during high-risk movements. With the inclusion of these neuromuscular and biomechanical performance measurements, machine learning models could be utilized to identify risk factors that are more injury-specific, which would be advantageous to modeling and understanding injury risk profiles. Furthermore, to the best of our knowledge, no previous research has used the LASSO regression for the predictive modeling of lower extremity injury risk in youth elite soccer players. These contemporary tools, along with the use of resampling methods to assess the models' predictive power, may overcome the limitations inherent to the current evidence and facilitate the construction of robust, interpretable, and generalizable models to predict lower extremity injuries.

Therefore, the purpose of this study was to (1) identify neuromuscular and biomechanical injury risk factors in elite youth soccer players and (2) assess the predictive ability of a machine learning approach using a LASSO model.

2 | MATERIALS AND METHODS

2.1 | Study design, participants, and injury data collection

This was a prospective cohort study. Six teams (under 16, under 17, and under 19) from the youth academies of two professional German soccer clubs were contacted and invited to participate. Three teams (under 16, under 17, and under 19) volunteered to take part in the study. Sixty-two elite male youth soccer players (age: 17.2 ± 1.1 years; height: 179 ± 8 cm; mass: 70.4 ± 9.2 kg) gave their written informed consent to participate. Players who met the inclusion criteria were tested via laboratory-based injury risk screening at the beginning of the preparation period for the 2018/2019 season. Injuries were documented by a standardized injury form throughout the 10 months of the 2018/2019 season starting after the initial testing at the beginning of the preparation period. The injury data collection process followed the procedures outlined in a consensus statement on injury definitions and data collection procedures in studies of soccer injuries.[20] Only time-loss non-contact injuries were counted, meaning that a player was missing completely in training or match play. Contact injuries, illnesses, and injuries to the upper body were excluded because of their unpredictable nature compared to non-contact injury mechanisms. Further information about the inclusion and exclusion criteria, as well as the injury data collection process, has been described in detail elsewhere.[21] The Ethics Committee of TU Dortmund University confirmed that the requirements of the Declaration of Helsinki were met.

2.2 | Laboratory-based injury risk screening

Prior to testing, all players performed a standardized warm-up, including 5 min cycling on a bike ergometer, followed by movement preparation and plyometric exercises. The laboratory-based injury risk screening battery consisted of a 3D motion analysis, postural control testing, and strength testing. After the warm-up, the players underwent these tests in a standardized order to avoid neuromuscular fatigue in the lower extremities and the trunk, particularly while undergoing 3D motion analysis and postural control testing, and to ensure equal test conditions for all players (see Figure 1). The procedures for data measurement and processing for the postural control testing, the strength testing and the 3D motion analysis have already been described in detail in previous studies.[21,22]

2.2.1 | 3D motion analysis

The players completed a single-leg drop landing task (SLDL) (30-cm box) and an unanticipated side-step cutting task (USSC) (45° cutting angle) with both legs. All landing and cutting tasks were performed on two force plates (AMTI Inc.) measuring 0.9 × 0.6 m and sampling at 1000 Hz. Each participant's lower body motion was captured using a 3D motion capture system consisting of 12 infrared cameras (120 Hz, Qualisys), which was time-synchronized to the force plates and a lower-body marker set of 40 markers.[23] The procedure for data processing was carried out in accordance with proven standards.[24,25] Hip and knee flexion-extension, abduction-adduction and internal-external rotation angles, ankle dorsiflexion-plantarflexion, eversion-inversion, and internal-external rotation angles were

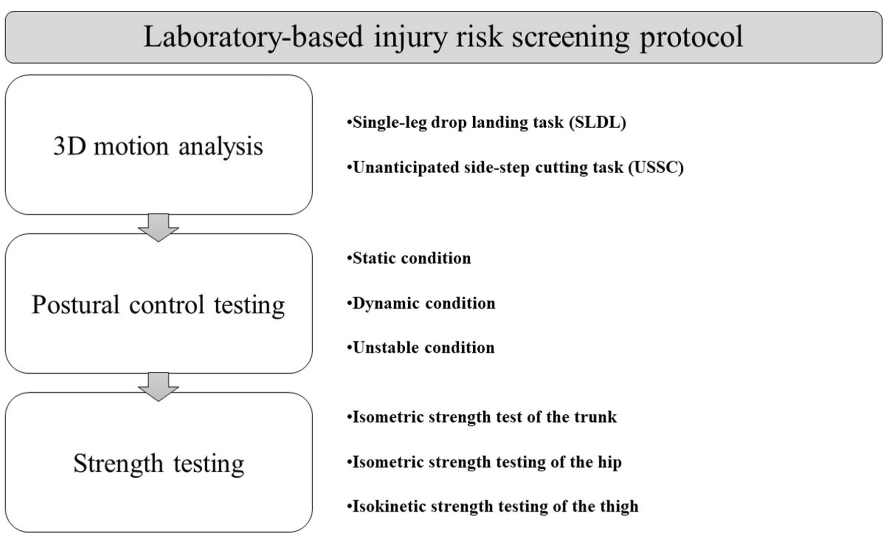

FIGURE 1 Three-step laboratory-based injury risk screening protocol in male youth elite soccer.

calculated as kinematic variables. The measured kinetics implied hip and knee flexion-extension, abduction-adduction, and internal-external rotation moments; ankle dorsiflexion-plantarflexion, eversion-inversion, and internal-external rotation moments; and vertical GRF (vGRF). Joint moments were expressed as the external moment applied to the joint. Body mass was used to normalize all kinetic parameters. Kinematic parameters were determined at two points during the stance phase: initial contact (IC) and the instance of peak value (PEAK), whereas kinetic parameters were determined at PEAK only. IC was defined as the first instance of the ground contact phase, whereas PEAK was defined as the peak value within the first 100 ms after IC.[26]

2.2.2 | Postural control testing

Postural control was assessed using three unilateral exercises under different conditions (see Figure 1). A single-leg stance test was used to analyze static postural control and postural control under unstable conditions. In each test, the players were instructed to start on the right leg and maintain balance for 10 s (static condition) or 20 s (unstable condition) in a static position with their eyes open. The hands were attached to the hips, and the swinging leg was flexed 90° in the hip, knee, and ankle to minimize contributions from the contralateral leg.[27] Dynamic postural control was analyzed using the Dynamic Postural Stability Index (DPSI) devised by Wikstrom et al,[28] which was modified to simulate soccer-specific movement.[21] The players were asked to jump off with both legs and perform an imaginary header before landing with one leg on a force plate (AMTI Inc.). Each player was to land first on the right leg, stabilize as quickly as possible, and balance for 3 s, with hands on the hips and looking straight ahead with the eyes open. For postural control under static and dynamic conditions, ground reaction force data during static and dynamic conditions were collected at a sampling frequency of 1000 Hz. Postural control under unstable conditions was measured while standing on a multi-axial free-swinging platform (Posturomed, Haider Bioswing). Three trials were performed on each leg for each testing condition. The lowest center of pressure (COP) sway was used as an outcome measure for postural control under static conditions, while the lowest DPSI score was determined for postural control under dynamic conditions, and the shortest path of the platform for postural control under unstable conditions. The lowest values were used because players with a low value tend to have better postural control.[21]

2.2.3 | Strength testing

The players also performed strength testing of the trunk, hip, and thigh muscles. Isometric trunk flexion, extension, lateral flexion, and transversal rotation were measured in all anatomic movement planes (sagittal–frontal–transverse) with the Pegasus 3-D system (Biofeedback Motor Control GmbH). Isometric hip adduction and abduction and the tests of concentric and eccentric knee extension and flexion were performed with an IsoMed 2000 isokinetic dynamometer (D&R Ferstl GmbH). For the isometric tests, the players performed three maximal voluntary isometric contractions of 5 s with a 1-min rest period in between.[29] The highest isometric peak torque normalized to body mass ($N \cdot m\,kg^{-1}$) for trunk flexion, trunk extension, trunk lateral flexion (right and left), trunk transversal rotation (right and left), hip abduction, and hip adduction was determined. For the isokinetic knee tests, the players performed two submaximal repetitions (70% of MVC), followed by three maximal repetitions at 60°/s across a range of motion of 10–90° with a 1-min rest between each testing condition and each contraction mode. The highest gravity corrected peak torque normalized to body mass ($N \cdot m\,kg^{-1}$) obtained for knee extension and knee flexion in both contraction modes (concentric and eccentric) was determined during the isokinetic phase.[21]

2.3 | Data analysis

Descriptive data are presented as means ± standard deviations. The primary statistical analysis in the present work is based on a LASSO regression model.[14,15] LASSO is an approach for regression analysis that performs both variable selection and regularization to enhance the predictive accuracy and interpretability of the statistical model it produces.[4] More specifically, the LASSO is well-known to be beneficial in terms of the estimates' mean squared error and, consequently, also in predictive settings, as the penalty parameter λ is tuned based on the prediction of unseen test data, for example, via K-fold cross validation (CV). In the following, we fit a LASSO-penalized logit model for the binary outcome *injury* (yes = 1/no = 0) and incorporating all available covariates (see Tables 1 and 2; main effects only; not interaction effects) with linear effects using the R package glmnet based on all the covariates introduced in Tables 1 and 2. The optimal penalty parameter λ was derived via 15-fold CV using the cv.glmnet function. The corresponding (pseudo) model equation is given by

TABLE 1 Descriptive statistics (mean ± standard deviation) for neuromuscular performance parameters of all players investigated according to their injury state (injured/non-injured).

	Performance parameter	All players ($n = 56$)	Injured players ($n = 22$)	Non-injured players ($n = 34$)
Postural control				
Static	COP sway (cm)	119.2 ± 24.1	129.4 ± 25.8	112.4 ± 20.5
Dynamic	DPSI	4.59 ± 0.93	4.50 ± 1.03	4.65 ± 0.87
Unstable	Path of platform (mm)	0.39 ± 0.19	0.44 ± 0.22	0.36 ± 0.16
Strength				
Trunk (isometric)	Flex ($N \cdot m\,kg^{-1}$)	2.35 ± 0.48	2.24 ± 0.48	2.42 ± 0.48
	Ext ($N \cdot m\,kg^{-1}$)	4.96 ± 1.06	4.86 ± 0.95	5.02 ± 1.13
	Flex + Ext ($N \cdot m\,kg^{-1}$)	3.65 ± 0.67	3.55 ± 0.63	3.72 ± 0.7
	Flex/Ext	0.49 ± 0.12	0.47 ± 0.12	0.50 ± 0.12
	LatFlex ($N \cdot m\,kg^{-1}$)	2.41 ± 0.49	2.32 ± 0.45	2.48 ± 0.52
	$LatFlex_r/LatFlex_l$	0.98 ± 0.17	0.98 ± 0.12	0.98 ± 0.20
	TransRot ($N \cdot m\,kg^{-1}$)	1.99 ± 0.35	1.94 ± 0.35	2.02 ± 0.34
	$TransRot_r/TransRot_l$	1.02 ± 0.16	1.05 ± 0.18	0.99 ± 0.14
	Core Score ($N \cdot m\,kg^{-1}$)	16.1 ± 2.57	15.6 ± 2.47	16.4 ± 2.62
Hip (isometric)	ABD ($N \cdot m\,kg^{-1}$)	1.92 ± 0.32	1.90 ± 0.31	1.93 ± 0.33
	ADD ($N \cdot m\,kg^{-1}$)	2.07 ± 0.53	1.99 ± 0.45	2.13 ± 0.57
	ABD/ADD	0.96 ± 0.19	0.98 ± 0.16	0.94 ± 0.21
Knee (isokinetic)	Qcon ($N \cdot m\,kg^{-1}$)	3.08 ± 0.47	2.86 ± 0.46	3.23 ± 0.42
	$Qcon_l/Qcon_r$	1.01 ± 0.14	1.04 ± 0.17	0.98 ± 0.11
	Hcon ($N \cdot m\,kg^{-1}$)	1.67 ± 0.25	1.57 ± 0.20	1.73 ± 0.27
	$Hcon_l/Hcon_r$	0.97 ± 0.11	0.99 ± 0.1	0.96 ± 0.12
	Qecc ($N \cdot m\,kg^{-1}$)	3.66 ± 0.70	3.40 ± 0.69	3.83 ± 0.66
	$Qecc_l/Qecc_r$	1.02 ± 0.18	1.05 ± 0.19	0.99 ± 0.18
	Hecc ($N \cdot m\,kg^{-1}$)	2.13 ± 0.41	1.99 ± 0.32	2.22 ± 0.45
	$Hecc_l/Hecc_r$	1.01 ± 0.14	1.02 ± 0.14	1.01 ± 0.14
	Conventional knee ratio: Hcon/Qcon	0.55 ± 0.07	0.56 ± 0.07	0.54 ± 0.07
	Functional knee ratio: Hecc/Qcon	0.69 ± 0.12	0.70 ± 0.11	0.69 ± 0.12

Abbreviations: ABD, hip abduction; ABD/ADD, ratio between hip abduction and hip adduction; ADD, hip adduction; Conventional knee ratio, ratio between knee flexion concentric and knee extension concentric; Core Score, sum of trunk flexion, trunk extension, trunk lateral flexion right, trunk lateral flexion left, trunk transversal rotation right and trunk transversal rotation left; COP, center of pressure; DPSI, Dynamic Postural Stability Index; Ext, trunk extension; Flex + Ext, sum of trunk flexion and trunk extension; Flex, trunk flexion; Flex/Ext, ratio between trunk flexion and trunk extension; Functional knee ratio, ratio between knee flexion eccentric and knee extension concentric; Hcon, knee flexion concentric; Hecc, knee flexion eccentric; LatFlex, trunk lateral flexion; LatFlexl, trunk lateral flexion left; LatFlexr, trunk lateral flexion right; LatFlexr/LatFlexl, ratio between trunk lateral flexion right and trunk lateral flexion left; Qcon, knee extension concentric; Qecc, knee extension eccentric; TransRot, trunk transversal rotation; TransRotl, trunk transversal rotation left; TransRotr, trunk transversal rotation right; TransRotr/TransRotl, ratio between trunk transversal rotation right and trunk transversal rotation left.

$$\text{injury (yes/no)} \sim \text{icept} + \text{biomechanical features} + \text{neuromuscular features} + \text{sociodemographic features}.$$

Note that for a proper usage of the LASSO penalization technique, all covariates have to be standardized to make the penalty comparable.

Along with the standard LASSO model, we have also introduced two extensions. First, in addition to the optimal penalty parameter λ, which minimizes the CV error, we have also considered a slightly weaker penalty strength λ_{1se} that gives the least regularized model, such that the CV error is within one standard error of the minimum (see Figure 2). Second, in addition to the standard LASSO estimates, we have also calculated the so-called post-LASSO estimates, which are based on the idea of the relaxed LASSO.[30] The

TABLE 2 Descriptive statistics (mean ± standard deviation) for biomechanical performance parameters during the SLDL and USSC of all players investigated according to their injury state (injured/non-injured).

	All players (n = 56)	Injured players (n = 22)	Non-injured players (n = 34)
Joint kinematics at IC (°) and PEAK (°) during SLDL			
Ankle			
Plantarflexion(+)/Dorsalflexion(−) IC	24.2 ± 5.3	24.2 ± 5.1	24.1 ± 5.5
Plantarflexion(+)/Dorsalflexion(−) PEAK	26.1 ± 6.1	26.9 ± 6.4	25.6 ± 6.0
Eversion(+)/Inversion(−) IC	−8.2 ± 3.8	−8.2 ± 4.2	−8.3 ± 3.5
Eversion(+)/Inversion(−) PEAK	−9.4 ± 3.9	−9.6 ± 4.2	−9.2 ± 3.7
External Rotation(+)/Internal Rotation(−) IC	−1.8 ± 3.9	−1.5 ± 4.0	−2.0 ± 3.9
External Rotation(+)/Internal Rotation(−) PEAK	−3.2 ± 4.5	−2.9 ± 4.6	−3.4 ± 4.4
Knee			
Flexion(+)/Extension(−) IC	15.1 ± 5.0	15.0 ± 4.8	15.1 ± 5.2
Flexion(+)/Extension(−) PEAK	52.3 ± 4.6	53.0 ± 4.3	51.8 ± 4.8
Adduction(+)/Abduction(−) IC	2.0 ± 2.9	1.6 ± 3.0	2.2 ± 2.9
Adduction(+)/Abduction(−) PEAK	−0.9 ± 4.0	−2.0 ± 4.6	−0.1 ± 3.4
External Rotation(+)/Internal Rotation(−) IC	2.0 ± 4.0	1.5 ± 4.2	2.4 ± 3.8
External Rotation(+)/Internal Rotation(−) PEAK	−10.0 ± 4.4	−9.4 ± 4.6	−10.4 ± 4.3
Hip			
Flexion(+)/Extension(−) IC	30.0 ± 6.9	30.1 ± 7.5	30.0 ± 6.6
Flexion(+)/Extension(−) PEAK	44.4 ± 8.6	44.4 ± 9.8	44.4 ± 7.9
Adduction(+)/Abduction(−) IC	−9.6 ± 2.8	−10.0 ± 2.7	−9.4 ± 3.0
Adduction(+)/Abduction(−) PEAK	0.3 ± 3.3	0.4 ± 3.5	0.2 ± 3.2
External Rotation(+)/Internal Rotation(−) IC	4.8 ± 4.5	4.8 ± 5.6	4.8 ± 3.8
External Rotation(+)/Internal Rotation(−) PEAK	−0.6 ± 4.2	−1.1 ± 4.9	−0.2 ± 3.7
Joint kinematics at IC (°) and PEAK (°) during USCC			
Ankle			
Plantarflexion(+)/Dorsalflexion(−) IC	−2.9 ± 9.2	−1.7 ± 11.1	−3.8 ± 7.7
Plantarflexion(+)/Dorsalflexion(−) PEAK	3.3 ± 6.5	4.7 ± 6.9	2.4 ± 6.1
Eversion(+)/Inversion(−) IC	−12.8 ± 5.8	−13.0 ± 6.0	−12.7 ± 5.7
Eversion(+)/Inversion(−) PEAK	−22.0 ± 4.9	−22.2 ± 5.3	−21.9 ± 4.7
External Rotation(+)/Internal Rotation(−) IC	3.1 ± 5.1	3.1 ± 5.9	3.1 ± 4.5
External Rotation(+)/Internal Rotation(−) PEAK	−0.2 ± 5.2	0.9 ± 5.4	−1.0 ± 5.0
Knee			
Flexion(+)/Extension(−) IC	42.8 ± 12.1	38.6 ± 12.3	45.6 ± 11.2
Flexion(+)/Extension(−) PEAK	56.6 ± 5.4	55.5 ± 6.0	57.4 ± 4.8
Adduction(+)/Abduction(−) IC	−1.1 ± 4.3	−1.1 ± 4.0	−1.1 ± 4.6
Adduction(+)/Abduction(−) PEAK	−6.0 ± 4.7	−6.2 ± 4.7	−5.9 ± 4.8
External Rotation(+)/Internal Rotation(−) IC	−7.1 ± 5.8	−6.1 ± 6.4	−7.8 ± 5.3
External Rotation(+)/Internal Rotation(−) PEAK	−12.6 ± 5.4	−11.5 ± 5.8	−13.4 ± 5.0
Hip			
Flexion(+)/Extension(−) IC	66.7 ± 8.9	64.0 ± 10.7	68.5 ± 7.0
Flexion(+)/Extension(−) PEAK	68.2 ± 8.6	65.6 ± 10.1	70.0 ± 7.0
Adduction(+)/Abduction(−) IC	−7.1 ± 4.4	−7.1 ± 4.5	−7.2 ± 4.3
Adduction(+)/Abduction(−) PEAK	−3.6 ± 4.6	−4.0 ± 4.7	−3.3 ± 4.5
External Rotation(+)/Internal Rotation(−) IC	−2.6 ± 4.7	−2.5 ± 5.9	−2.7 ± 3.7
External Rotation(+)/Internal Rotation(−) PEAK	6.1 ± 5.4	6.3 ± 7.1	6.0 ± 4.1

TABLE 2 (Continued)

	All players (n = 56)	Injured players (n = 22)	Non-injured players (n = 34)
PEAK joint moments (Nm/kg)			
Ankle			
Plantarflexion(+)/Dorsalflexion(−) SLDL	−0.1 ± 0.0	−0.1 ± 0.0	−0.1 ± 0.0
Plantarflexion(+)/Dorsalflexion(−) USSC	0.1 ± 0.2	0.2 ± 0.2	0.1 ± 0.2
Eversion(+)/Inversion(−) SLDL	−0.4 ± 0.1	−0.4 ± 0.1	−0.4 ± 0.1
Eversion(+)/Inversion(−) USSC	−0.3 ± 0.1	−0.3 ± 0.1	−0.3 ± 0.1
External Rotation(+)/Internal Rotation(−) SLDL	−0.4 ± 0.1	−0.4 ± 0.1	−0.4 ± 0.1
External Rotation(+)/Internal Rotation(−) USSC	−0.1 ± 0.1	−0.1 ± 0.1	−0.1 ± 0.1
Knee			
Flexion(+)/Extension(−) SLDL	2.5 ± 0.4	2.5 ± 0.4	2.5 ± 0.4
Flexion(+)/Extension(−) USSC	2.2 ± 0.5	2.3 ± 0.5	2.2 ± 0.5
Adduction(+)/Abduction(−) SLDL	−0.7 ± 0.3	−0.7 ± 0.3	−0.7 ± 0.3
Adduction(+)/Abduction(−) USSC	−0.9 ± 0.2	−0.8 ± 0.2	−0.9 ± 0.2
External Rotation(+)/Internal Rotation(−) SLDL	−0.1 ± 0.1	−0.1 ± 0.1	−0.1 ± 0.0
External Rotation(+)/Internal Rotation(−) USSC	−0.3 ± 0.2	−0.3 ± 0.3	−0.3 ± 0.2
Hip			
Flexion(+)/Extension(−) SLDL	3.6 ± 0.6	3.7 ± 0.6	3.6 ± 0.5
Flexion(+)/Extension(−) USSC	3.7 ± 0.7	3.6 ± 0.7	3.8 ± 0.6
Adduction(+)/Abduction(−) SLDL	1.0 ± 0.2	1.0 ± 0.2	1.0 ± 0.2
Adduction(+)/Abduction(−) USSC	0.6 ± 0.3	0.6 ± 0.3	0.6 ± 0.3
External Rotation(+)/Internal Rotation(−) SLDL	−0.6 ± 0.2	−0.7 ± 0.3	−0.6 ± 0.1
External Rotation(+)/Internal Rotation(−) USSC	−0.6 ± 0.2	−0.7 ± 0.3	−0.6 ± 0.1
PEAK vGRF (N/kg)			
vGRF SLDL	38.1 ± 3.9	37.7 ± 4.7	38.4 ± 3.4
vGRF USSC	20.9 ± 2.5	20.4 ± 2.4	21.2 ± 2.5

Abbreviations: °, degrees; IC, initial contact: first instance of ground contact phase; kg, kilogram; N, newton; Nm, newton meter; PEAK, peak value: peak value within the first 100 ms after IC; SLDL, single-leg drop landing; USSC, unanticipated side-step cutting; vGRF, vertical ground reaction force.

relaxed LASSO allows to use linear combinations between the (typically strongly penalized) standard LASSO estimates and the unregularized maximum likelihood (ML) estimator. For the special case of the post-LASSO, the LASSO is only used for variable selection, and then, a completely unpenalized ML regression is performed on the chosen subset of selected covariates. This technique was recently implemented in glmnet via the relax argument.

We investigated the predictive performance of all these models in a leave-one-out (LOO) prediction competition, in which each of the 57 observations was once left out, while the remaining 56 observations were used to train the models. The left-out observation is then predicted by all models, and predictive performance is evaluated by the area under the ROC curve (AUC),[8,31] as well as the predictive Bernoulli likelihood. Generally, the AUC could be classified as outstanding (0.90–1), excellent (0.80–0.89), acceptable (0.70–0.79), poor (0.51–0.69), or no discrimination (0.50).[32] Additionally, sensitivity and specificity were calculated and served as performance measures of the LASSO model. Sensitivity measures the proportion of injured players who were correctly predicted by the LASSO model as being injured, while specificity measures the proportion of non-injured players correctly predicted as such. For comparison, as a simple benchmark approach, a classifier that guesses the two outcomes based solely on their relative proportions has been included in the competition.

3 | RESULTS

3.1 | Epidemiology

In the present study, 62 elite youth soccer players (age: 17.2 ± 1.1 years; height: 179 ± 8 cm; weight: 70.4 ± 9.2 kg) were enrolled, while one player was injured twice. Because of technical problems, the data of six players could not be included in the analysis, which resulted in an attrition

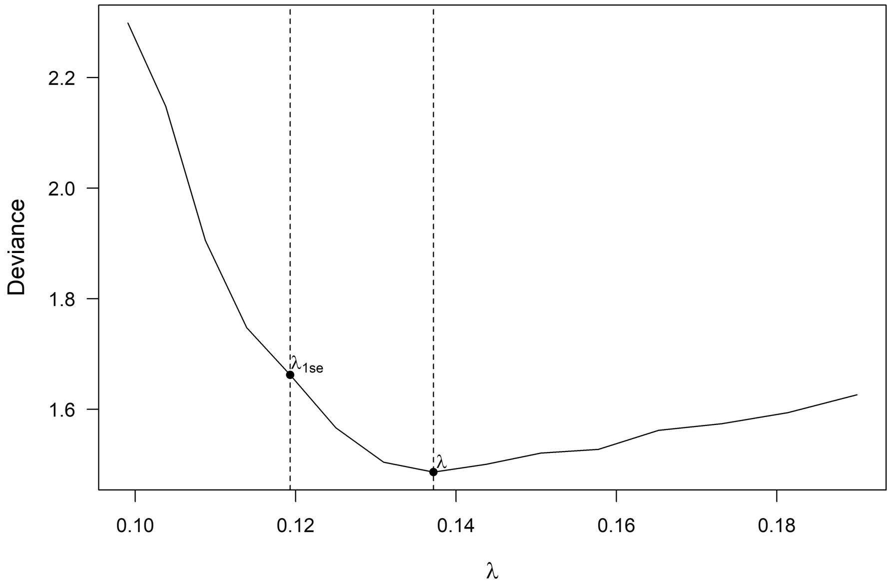

FIGURE 2 Deviance derived from the cross validation to optimize lambda.

rate of 9.6%. Over the 2018/2019 season, 23 non-contact injuries were registered, with 39% of the players having one or more. The overall non-contact injury incidence was 1.2/1000 h of total exposure time (0.5 injuries per 1000 h of training and 3.9 injuries per 1000 h of competition). The most frequently injured body parts were the ankle (36%) and thigh muscles (hamstrings: 18%; quadriceps: 18%), followed by the adductors (16%) and the knee joint (12%). Sprains (48%) and strains (39%) were the most common injury types. Table 1 presents the descriptive statistics for the neuromuscular parameters, whereas Table 2 presents the descriptive statistics for the biomechanical performance parameters.

3.2 | LASSO and predictive performance

The results of the LOO prediction competition are shown in Table 3. The two LASSO variants performed slightly better than the benchmark in terms of predictive likelihood. The best model in this regard was the post-LASSO one based on the weaker penalty strength, with a value of 0.58. It also showed the highest AUC of 0.63.

The post-LASSO model predicted 35% of the injured players as being injured, while 79% of the non-injured players were predicted by the model to remain as such. These numbers correspond to the highest sensitivity and the lowest specificity among the four fitted models. However, the model's specificity was only 12% lower than the highest value achieved among all models, while its sensitivity was 13% higher than the model with the second-highest value. Therefore, the post-LASSO λ_{1se} model was considered the best in the competition.

The post-LASSO λ_{1se} model identified concentric knee extensor peak torque, hip transversal plane moment in the SLDL and COP sway as the three most important predictors for injury in that order according to sizes of the estimated model coefficients (see Table 4). As these coefficients refer to the standardize covariate values, they are directly comparable.

4 | DISCUSSION

The development of a predictive model can improve our understanding of how physical performance parameters affect injury risk. Accordingly, this study sought to identify neuromuscular and biomechanical injury risk factors in elite youth soccer players and to assess their predictive abilities using a LASSO regression model.

The complexity of the interactions between various physical performance parameters and the number of confounding variables during the occurrence of an injury makes it difficult to predict the cause of lower extremity injuries using a suitable statistical model.[33] In addition, laboratory-based screening tests are time-consuming, reducing the opportunity of having an extensive, robust data set in an elite cohort, which will improve the ability of supervised learning techniques to detect patterns with more consistency as the number of injuries increases.[8] However, the measurement properties of laboratory-based screening tests are well established. Therefore, these measurement methods are considered the gold standard by which patterns that expose players to a higher risk of injury can be validly identified.

Considering these aspects, we were nevertheless able to build a predictive model with our approach using LASSO penalization. The concentric knee extensor peak torque, hip transversal plane moment in the SLDL, and

TABLE 3 Predictive performance measures derived from LOO.

Method	Sensitivity	Specificity	AUC	Predictive likelihood
Simple classifier	0.57	0.56	—	0.51
LASSO λ	0.09	0.91	0.52	0.52
LASSO λ_{1se}	0.17	0.82	0.47	0.51
Post-LASSO λ	0.22	0.82	0.56	0.55
Post-LASSO λ_{1se}	0.35	0.79	0.63	0.58

Abbreviations: AUC, area under the curve; LASSO, least absolute shrinkage and selection operator; LOO, leave-one-out.

TABLE 4 Results of the fitted post-LASSO λ_{1se} model on full data; coefficients correspond to standardized features.

Model parameter	Model coefficient estimates $\hat{\beta}$
(Intercept)	−0.40
COP sway	0.73
concentric knee extensor peak torque	−0.97
hip transversal plane moment in the SLDL	−0.90

Abbreviations: COP, center of pressure; LASSO, least absolute shrinkage and selection operator; SLDL, single-leg drop landing.

the COP sway were identified by the post-LASSO λ_{1se} model in that order and served as injury predictors. The predictive performance measures of the post-LASSO λ_{1se} model indicated a good ability to identify players who did not get injured (specificity = 79%) but not those who did get injured (sensitivity = 35%). Furthermore, the model was able to predict injury outcomes with a likelihood of 58% and an AUC of 0.63. The predictive performance, especially the AUC value obtained, is in line with the results of previous studies on injury prediction in different populations[11,18,34] and shows the challenge of predicting sports injuries. Although the AUC is important, researchers should also consider the need for high sensitivity when designing models that identify players at risk of injury.[11,35] It is relatively easy to create models with high specificity due to an often imbalanced dataset, as a model tends to become more specific and less sensitive, since classifying the non-injured correctly has a greater impact on the overall classification rate. Interestingly, the highest sensitivity in the present study corresponded to the best of the four fitted models. Nonetheless, the sensitivity remained rather low.

4.1 | Injury predictors

The three most important variables for predicting the injury outcome are concentric knee extensor peak torque, hip transversal plane moment in the SLDL, and COP sway, suggesting the importance of neuromuscular and biomechanical performance measures in elite youth soccer. A lower concentric strength of the knee extensors ($\hat{\beta}$ =−0.97), a higher hip internal rotation moment in the SLDL ($\hat{\beta}$ =−0.90), and a higher COP sway ($\hat{\beta}$ = 0.73) increased the estimated risk of injury. This supports a body of literature that found relationships between these variables and the risk of injury.[36–39] One of the possible reasons why these variables were chosen for the final model may be related to injury location and type. A total of 87% of all non-contact injuries in the present study were sprains and strains, with the ankle and thigh muscles being the most frequently injured body parts. Moreover, these variables play a crucial role in the hazardous lower extremity movement patterns that soccer players perform. In particular, an injury ultimately occurs in these high-risk situations when tissue stress exceeds the tissue's maximal capacity. Therefore, altered biomechanical motion and reduced neuromuscular control of the lower extremity can lead to exceeding the stress tolerance for muscles as well as tendons and ligaments in the joint.[40] According to the joint by joint approach,[41] poor postural control in the ascending kinematic chain without muscular stabilization can have an impact on hip movement patterns, especially in single-leg movements.

4.2 | Predictive performance measures

The predictive performance measures of the best model in the present study were lower than those reported by previous studies especially on elite youth soccer players that investigated the ability to predict injuries using supervised learning techniques.[11,17] In detail, Rommers et al[17] identified anthropometric measures as the most important variables for predicting injury with reasonably high precision and accuracy (AUC = 0.85; sensitivity = 85%; specificity = 85%). Oliver et al[11] concluded that the best performing decision tree identified asymmetry in the single-leg countermovement jump, asymmetry in the 75% Hop, asymmetry in the Y-Balance Test, knee valgus angle assessed through the tuck jump test, and body size as the most frequent contributors (AUC = 0.66; sensitivity = 56%; specificity = 74%). One plausible explanation for the apparent discrepancy between previous results in elite youth soccer and ours could be attributed to the higher sample size in their studies (n = 734 participants[17]; n = 355 participants[11]). Having a large robust dataset has been shown to improve the ability of supervised learning techniques to detect patterns with more consistency as the number of injuries increases.[8] Another explanation could be the different age distributions in the samples. Rommers et al,[17] for example, examined players from the U10–U15 age group. In a similar vein, Rumpf and Cronin[42] recently undertook a review of the literature examining injury incidence data in 6- to 18-year-old soccer players. Their results suggest that there is a significant age effect in the incidence of soccer-related injuries during childhood and that children aged 13–15 years are most at risk.[42] Injury incidence has also been proven to be related to the level of maturity, especially around the period of peak height velocity.[43] This age range would coincide with progression through puberty and all of the associated developmental changes (e.g., rapid and asynchronous growth of the musculoskeletal system) that occur during this time.[44] Thus, it does not appear to be surprising that Rommers et al[17] identified anthropometric measures as the most important predictors with high accuracy. However, although these variables give practitioners an estimation of the players who are at increased risk of injury, they cannot be modified by preventive measures, which are of outstanding importance in injury prevention.[45]

Using the more injury-specific testing procedures proposed by Oliver et al[11] resulted in decreased predictive performance, which is in line with the results of the present study. Interestingly, the use of more injury-specific testing procedures resulted in the selection of neuromuscular and biomechanical variables, which is also similar to the results of the present study. This shows the importance of multifactorial approaches in assessing the risk of non-contact lower extremity injury. However, it should be highlighted that Oliver et al[11] were limited in their inclusion of the gold-standard measurements of muscle strength, postural stability, and biomechanics of the lower extremities during high-risk movements. Instead, they used field-based and low-cost screening tests, whose measurement properties and relationships to injury are currently limited, conflicting, or unknown.[19] At the same time, they investigated a sample that was more than six times larger than ours.

Results in other populations show contradictory results. A systematic review[10] found eight studies that reported appropriate to good performance of the prediction models. In detail, AUC values for predicting the outcome ranged between 0.64 and 0.87, and high values were found for accuracy (75%–82.9%), sensitivity (55.6%–94.5%), specificity (74.2%–87%), and precision (50%–85%).[10] Contrary, few studies reported low predictive performance measures, showing poor AUC (0.52–0.65) and low accuracy (52%) values.[10] However, a comparison with the results of the present study is limited, because most of the included studies were performed in Australian Football, Rugby, Baseball, and men's soccer. It is therefore difficult to make an appropriate comparison because either the load profile of the sports differs significantly from that of youth soccer or there are differences in the maturity level of the athletes. Additionally, analyzed risk factors included both modifiable and non-modifiable factors without using gold-standard measurements.[10]

Therefore, we conclude that with the inclusion of gold-standard neuromuscular and biomechanical performance measurements in large samples, supervised learning models could be utilized to identify which risk factors are more injury specific and which would be advantageous to modeling, as well as to improving our understanding of injury risk profiles.

4.3 | Limitations

The lack of predictive performance can be attributed to a number of factors. First, the data were collected only at the beginning of the preseason, and it remains unknown whether more frequent measurements would have improved the model's predictive performance. Moreover, injury prediction has been presented as a non-linear system, as athletes are dynamic systems[46]; however in the present analysis in the framework of LASSO regression, only linear effects could be investigated. Additionally, a single preseason evaluation of physical performance through a laboratory-based injury risk screening does not reflect an athlete's physical performance during the injury onset, as there is often a period of weeks to months between the preseason measurement and the occurrence of the injury. During this time, the underlying biomechanical or neuromuscular factors causing the injury may change or evolve for a multitude of reasons, especially in youth soccer players.[46] Therefore, as suggested by previous research,[47] the incorporation of daily internal and external workload data, which produce real-time, continuous data that are more accessible, in addition to biomechanical and neuromuscular risk factors, may improve the prediction of lower extremity injuries and the establishment of injury risk profiles. Initial results indicate that especially the daily use of

GPS tracking technology in professional soccer, in particular, leads to the early detection of an increased risk of injury.[48]

4.4 | Practical considerations

Developing a "practitioner-friendly" predictive model to profile the risk of sustaining a lower extremity injury in youth soccer would help minimize the risk of such injuries. Our final model considered three of the 89 variables that were initially measured in the present study. This may suggest that the range of laboratory-based injury screenings that are necessary to identify injury predictors is relatively manageable in soccer-specific settings due to the reduced time required for the testing procedures in the preseason examinations. However, this practical consideration must be interpreted with caution due to the poor predictive performance of the final model, which must be evaluated in larger samples.

The main categories of potential risk factors for non-contact lower extremity injuries (neuromuscular and biomechanical) are represented in the final model. This shows the importance of multifactorial approaches in assessing the risk of non-contact lower extremity injury. Moreover, identifying injury risk factors and assessing the individual risk of injury of each player enables customized injury prevention interventions to be provided as part of the player's daily training schedule.[21] Several reviews and meta-analyses dealing with the effects of multicomponent exercise prevention programs have demonstrated an effective reduction in injury risk.[49,50]

4.5 | Perspectives

To the best of our knowledge, this is the first study to use LASSO regression and laboratory-based injury risk screening for the predictive modeling of lower extremity injury risk in male youth elite soccer players. The three most important variables for predicting the injury outcome were found to be concentric knee extensor peak torque, hip transversal plane moment in the SLDL, and COP sway, suggesting the importance of neuromuscular and biomechanical performance measures in elite youth soccer. These preliminary results may have practical implications for future directions in injury risk screening and planning, as well as for the development of customized training programs to counteract intrinsic injury risk factors. However, the poor predictive performance of the final model confirms the challenge of predicting sports injuries, and the model must therefore be evaluated in larger samples.

ACKNOWLEDGEMENTS

The authors would like to thank the participating youth academies, coaches, and players for their collaboration and the test team for their contribution to the data collection. Open Access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no conflict of interest relevant to the content of this research.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

ORCID

Mathias Kolodziej https://orcid.org/0000-0002-6030-3578
Andreas Groll https://orcid.org/0000-0001-6787-9118
Tobias Alt https://orcid.org/0000-0002-4451-9512
Marcus Schmidt https://orcid.org/0000-0001-9417-544X

REFERENCES

1. Read PJ, Oliver JL, De Ste Croix MBA, Myer GD, Lloyd RS. The scientific foundations and associated injury risks of early soccer specialisation. *J Sports Sci*. 2016;34(24):2295-2302.
2. Pfirrmann D, Herbst M, Ingelfinger P, Simon P, Tug S. Analysis of injury incidences in male professional adult and elite youth soccer players: a systematic review. *J Athl Train*. 2016;51(5):410-424.
3. Hägglund 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(12):738-742.
4. Claudino JG, de Oliveira CD, de Souza TV, Serrão JC, Pereira ACM, Nassis GP. Current approaches to the use of artificial intelligence for injury risk assessment and performance prediction in team sports: a systematic review. *Sports Med Open*. 2019;5(1):1-12.
5. 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(21):1309-1314.
6. Bahr R, Holme I. Risk factors for sports injuries—a methodological approach. *Br J Sports Med*. 2003;37(5):384-392.
7. Read P, Oliver JL, De Ste Croix MBA, Myer GD, Lloyd RS. Injury risk factors in male youth soccer players. *Strength Cond J*. 2015;37(5):1-7.
8. Ruddy JD, Cormack SJ, Whiteley R, Williams MD, Timmins RG, Opar DA. Modeling the risk of team sport injuries: a narrative review of different statistical approaches. *Front Physiol*. 2019;10:829.

9. Fonseca ST, Souza TR, Verhagen E, et al. Sports injury forecasting and complexity: a synergetic approach. *Sports Med*. 2020;50(10):1757-1770.
10. Van Eetvelde H, Mendonça LD, Ley C, Seil R, Tischer T. Machine learning methods in sport injury prediction and prevention: a systematic review. *J Exp Orthop*. 2021;8(1):1-15.
11. 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(11):1044-1048.
12. Bühlmann P, Hothorn T. Boosting algorithms: regularization, prediction and model fitting. *Stat Sci*. 2007;22(4):477-505.
13. Breiman L. Random forests. *Mach Learn*. 2001;45(1):5-32.
14. Tibshirani R. Regression shrinkage and selection via the lasso. *J R Stat Soc B Methodol*. 1996;58(1):267-288.
15. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw*. 2010;33(1):1-22.
16. Ramadan A, Boss C, Choi J, et al. Selecting sensitive parameter subsets in dynamical models with application to biomechanical system identification. *J Biomech Eng*. 2018;140(7):745031-745038.
17. Rommers N, Rössler R, Verhagen E, et al. A machine learning approach to assess injury risk in elite youth football players. *Med Sci Sports Exerc*. 2020;52(8):1745-1751.
18. López-Valenciano A, Ayala F, Puerta JM, et al. A preventive model for muscle injuries: a novel approach based on learning algorithms. *Med Sci Sports Exerc*. 2018;50(5):915-927.
19. Hegedus E, McDonough SM, Bleakley C, Baxter D, Cook C. Clinician-friendly lower extremity physical performance tests in athletes: a systematic review of measurement properties and correlation with injury. Part 2—the tests for the hip, thigh, foot and ankle including the star excursion balance test. *Br J Sports Med*. 2015;49(10):649-656.
20. Fuller CW, Ekstrand J, Junge A, et al. Consensus statement on injury definitions and data collection procedures in studies of football (soccer) injuries. *Br J Sports Med*. 2006;40(3):193-201.
21. Kolodziej M, Nolte K, Schmidt M, Alt T, Jaitner T. Identification of neuromuscular performance parameters as risk factors of non-contact injuries in male elite youth soccer players: a preliminary study on 62 players with 25 non-contact injuries. *Front Sports Act Living*. 2021;3(295):615330.
22. Kolodziej M, Willwacher S, Nolte K, Schmidt M, Jaitner T. Biomechanical risk factors of injury-related single-leg movements in male elite youth soccer players. *Biomechanics*. 2022;2(2):281-300.
23. Schmidt M, Nolte K, Terschluse B, Jaitner T. Fatigue impairs kinematics but not kinetics of landing and cutting movements in elite youth female handball. *ISBS Proc Arch*. 2020;38(1):176.
24. Willwacher S, Konig M, Braunstein B, Goldmann JP, Bruggemann GP. The gearing function of running shoe longitudinal bending stiffness. *Gait Posture*. 2014;40(3):386-390.
25. Willwacher S, Kurz M, Menne C, Schrödter E, Brüggemann G-P. Biomechanical response to altered footwear longitudinal bending stiffness in the early acceleration phase of sprinting. *Footwear Sci*. 2016;8(2):99-108.
26. Kristianslund E, Krosshaug T. Comparison of drop jumps and sport-specific sidestep cutting: implications for anterior cruciate ligament injury risk screening. *Am J Sports Med*. 2013;41(3):684-688.
27. Frisch A, Urhausen A, Seil R, Croisier J-L, Windal T, Theisen D. Association between preseason functional tests and injuries in youth football: a prospective follow-up. *Scand J Med Sci Sports*. 2011;21(6):e468-e476.
28. Wikstrom EA, Tillman MD, Smith AN, Borsa PA. A new force-plate technology measure of dynamic postural stability: the dynamic postural stability index. *J Athl Train*. 2005;40(4):305-309.
29. Meyer C, Corten K, Wesseling M, et al. Test-retest reliability of innovated strength tests for hip muscles. *PLoS One*. 2013;8(11):e81149.
30. Meinshausen N. Relaxed lasso. *Comput Stat Data Anal*. 2007;52(1):374-393.
31. Fawcett T. An introduction to ROC analysis. *Pattern Recognit Lett*. 2006;27(8):861-874.
32. Hosmer DW Jr, Lemeshow S, Sturdivant RX. *Applied Logistic Regression*. Vol 398. John Wiley & Sons; 2013.
33. Henriquez M, Sumner J, Faherty M, Sell T, Bent B. Machine learning to predict lower extremity musculoskeletal injury risk in student athletes. *Front Sports Act Living*. 2020;2:2.
34. Ruddy J, Shield A, Maniar N, et al. Predictive modeling of hamstring strain injuries in elite Australian footballers. *Med Sci Sports Exerc*. 2018;50(5):906-914.
35. Florkowski CM. Sensitivity, specificity, receiver-operating characteristic (ROC) curves and likelihood ratios: communicating the performance of diagnostic tests. *Clin Biochem Rev*. 2008;29(Suppl 1):S83.
36. van Dyk N, Bahr R, Whiteley R, et al. Hamstring and quadriceps isokinetic strength deficits are weak risk factors for hamstring strain injuries: a 4-year cohort study. *Am J Sports Med*. 2016;44(7):1789-1795.
37. Tropp H, Ekstrand J, Gillquist J. Stabilometry in functional instability of the ankle and its value in predicting injury. *Med Sci Sports Exerc*. 1984;16(1):64-66.
38. McLean SG, Huang X, Van Den Bogert AJ. Association between lower extremity posture at contact and peak knee valgus moment during sidestepping: implications for ACL injury. *Clin Biomech*. 2005;20(8):863-870.
39. Freckleton G, Pizzari T. Risk factors for hamstring muscle strain injury in sport: a systematic review and meta-analysis. *Br J Sports Med*. 2012;4:2012-2358.
40. Read PJ, Oliver JL, De Ste Croix MB, Myer GD, Lloyd RS. Neuromuscular risk factors for knee and ankle ligament injuries in male youth soccer players. *Sports Med*. 2016;46(8):1059-1066.
41. Cook G. *Movement: Functional Movement Systems: Screening, Assessment, Corrective Strategies*. On Target Publications; 2010.
42. Rumpf MC, Cronin J. Injury incidence, body site, and severity in soccer players aged 6–18 years: implications for injury prevention. *Strength Cond J*. 2012;34(1):20-31.
43. Van der Sluis A, Elferink-Gemser M, Brink M, Visscher C. Importance of peak height velocity timing in terms of injuries in talented soccer players. *Int J Sports Med*. 2015;36(4):327-332.
44. Williams C, Wood L. Growth and maturation in childhood. In: De Ste Croix M, Korff T, eds. *Paediatric Biomechanics and Motor Control: Theory and Application*. Routledge; 2012:3-26.
45. van Mechelen W, Hlobil H, Kemper HC. Incidence, severity, aetiology and prevention of sports injuries. A review of concepts. *Sports Med*. 1992;14(2):82-99.
46. Stern BD, Hegedus EJ, Lai Y-C. Injury prediction as a nonlinear system. *Phys Ther Sport*. 2020;41:43-48.
47. Duhig S, Shield AJ, Opar D, Gabbett TJ, Ferguson C, Williams M. Effect of high-speed running on hamstring strain injury risk. *Br J Sports Med*. 2016;50(24):1536-1540.

48. Rossi A, Pappalardo L, Cintia P, Iaia FM, Fernández J, Medina D. Effective injury forecasting in soccer with GPS training data and machine learning. *PLoS One*. 2018;13(7):e0201264.
49. Al Attar WSA, Alshehri MA. A meta-analysis of meta-analyses of the effectiveness of FIFA injury prevention programs in soccer. *Scand J Med Sci Sports*. 2019;29(12):1846-1855.
50. Owoeye OB, VanderWey MJ, Pike I. Reducing injuries in soccer (football): an umbrella review of best evidence across the epidemiological framework for prevention. *Sports Med Open*. 2020;6(1):1-8.

How to cite this article: Kolodziej M, Groll A, Nolte K, et al. Predictive modeling of lower extremity injury risk in male elite youth soccer players using least absolute shrinkage and selection operator regression. *Scand J Med Sci Sports*. 2023;33:1021-1033. doi:10.1111/sms.14322