sports

MDPI

Article

Multidomain Physical Performance Profiling Discriminates Amateur from Professional Soccer Players: An Explainable Machine Learning Analysis

Gökhan Deliceoğlu [1], Fatih Bektaş [2], Adeviye Gamze Akın [3], Musab Çağın [1], Hüsnü Kocaman [1] and Erkan Tortu [4,*]

[1] Department of Coaching Education, Faculty of Sport Sciences, Gazi University, 06560 Ankara, Türkiye; gokhandeliceoglu@gazi.edu.tr (G.D.)
[2] Department of Recreation Education, Faculty of Sport Sciences, Trabzon University, 61335 Trabzon, Türkiye
[3] Department of Biostatistics, Faculty of Medicine, Hacettepe University, 06100 Ankara, Türkiye
[4] Department of Coaching Education, Faculty of Sport Sciences, Trabzon University, 61335 Trabzon, Türkiye
* Correspondence: erkantortu@trabzon.edu.tr

Abstract

The physical qualities that separate amateur from professional soccer players are usually studied one domain at a time. This study examined a multidomain physical performance battery with explainable machine learning. Cross-sectional data from 201 male players (101 amateur, 100 professional) were reduced from 45 variables to 24 curated features spanning anthropometry/training history, isokinetic strength, vertical jump, Wingate anaerobic power, reactive agility and sprint. Five classifiers were tuned with nested cross-validation and examined with resampling, permutation, SHapley Additive exPlanations and domain-restricted analyses. All variables except the hamstring:quadriceps ratios, countermovement–squat jump difference and stretch-shortening-cycle contribution differed between groups after false-discovery-rate correction (Cohen's d = 0.49–2.10). All classifiers achieved near-ceiling discrimination (test area under the receiver operating characteristic curve [AUC] 0.998–1.000; nested cross-validated AUC 0.994–0.999; label-permutation test of the logistic-regression pipeline $p = 0.001$), and the unsupervised composite index reached AUC = 0.969. Training age, jump height, 30 m sprint time and reactive agility contributed most, whereas body-mass-normalised isokinetic torque discriminated least (AUC 0.77–0.80); jump-plus-Wingate and sprint-plus-agility subsets alone reached AUC 0.95–0.98. Discriminatory information was distributed across domains rather than concentrated in isokinetic strength. These cross-sectional, cohort-specific findings describe established groups and do not support individual talent selection without external validation.

Keywords: isokinetic strength; sprint performance; change-of-direction agility; vertical jump; Wingate test; multicollinearity; explainable artificial intelligence; SHAP; talent identification; machine learning

Check for updates

Academic Editor: César Cavinato Cal Abad

Received: 21 October 2026
Revised: 6 November 2026
Accepted: 11 November 2026
Published: 13 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Soccer is an intermittent, high-intensity team sport in which brief but decisive actions, such as sprints, accelerations, changes of direction, jumps and duels, are superimposed on prolonged submaximal activity [1]. The quality of these explosive actions depends heavily on the force and power the lower limbs can generate, and in particular on those values expressed relative to body mass. Because relative strength and power reflect an athlete's

capacity to accelerate their own mass, they are generally stronger correlates of competitive standard than absolute values [2,3].

Isokinetic dynamometry provides a reliable assessment of knee flexor and extensor peak torque at fixed angular velocities. Normalised to body mass (PT/W), these values have often been used to characterise the neuromuscular profile associated with competitive level, from subelite to professional players [4–6]; by playing position [7–9]; and in relation to speed, agility and explosive power [10]. A closely related index, the hamstring:quadriceps (H:Q) ratio, is widely reported alongside peak torque, but the literature ties it more consistently to hamstring-strain injury risk than to competitive level per se [11–15]. Across professional, elite-academy and amateur players, H:Q differences were only small to moderate, and the ratio differentiated competitive levels poorly [6].

Complementary information on explosive and reactive lower-limb capacity comes from vertical-jump testing. Countermovement jump (CMJ) and squat jump (SJ) height, their derived power outputs, and elastic-energy indices such as the eccentric utilisation ratio (EUR; the CMJ:SJ height ratio [16]) and the stretch-shortening-cycle contribution (SSC%) are used to characterise the muscle–tendon system's ability to store and reuse elastic energy during rapid stretch-shortening actions [17–20]. Reactive strength and leg stiffness add to this picture [21,22]. CMJ/SJ differences have been reported between professional and amateur players [5,23], although no level-related differences were evident among elite male national-team players [24].

Sprint and change-of-direction performance form a further, largely independent domain. Most in-game sprints in soccer are short (<20–30 m) and rarely reach maximal velocity, so early acceleration capacity is particularly important [25,26]; it is also closely related to lower-limb power output [2,27]. Short-split sprint times, together with reactive agility and turning speed, have been used to differentiate competitive levels and to inform talent identification [28–33]; national-team and first-division players have sprinted 1–4% faster than lower-division and junior players [34].

Most existing amateur-versus-professional comparisons examine these domains largely in isolation, typically relying on isokinetic strength and a small number of jump indices [4–6,23,33]. How these domains perform when analysed jointly, and how a broader multidomain battery compares with them, therefore remains unclear. Professional–amateur differences are also evident in intermittent endurance and repeated-sprint ability despite similar maximal oxygen uptake [35,36]. Quantifying which qualities separate established professionals from amateurs provides benchmarks for players aiming to step up a level, but such cross-sectional separation should not be confused with prospective talent prediction, for which physical tests have shown limited prognostic value [28,37–39].

Machine learning and explainable artificial intelligence (AI) methods are increasingly applied to soccer performance and injury data because they can model the joint discriminative structure of many correlated variables simultaneously [40–47]. The same capacity creates a problem: when predictors are collinear, coefficient magnitude in linear models, and even split-based importance in tree ensembles, can misrepresent which variables drive a prediction [48–54]. SHapley Additive exPlanations (SHAP) [55] offer a theoretically grounded, model-agnostic alternative, but their attributions are themselves sensitive to dependence among features [56,57], and they have rarely been cross-checked against classical univariate diagnostics in this literature.

Because between-level differences in individual physical qualities are already well documented [4–6,23,33–36], the principal contribution of the present study is not to show again that professionals outperform amateurs on single tests but (a) to quantify how much separation a jointly modelled multidomain battery provides and how strongly each domain discriminates when modelled on its own and (b) to provide methodological guidance

on handling algebraically derived variables and correlated predictors when interpreting machine-learning feature importance. The present study addresses this gap in four ways: it (i) incorporates the full available physical performance battery for this cohort (anthropometry and training history, isokinetic strength [including H:Q ratio], vertical-jump and elastic-energy indices, Wingate anaerobic power, reactive agility, and linear sprint) rather than a single domain; (ii) applies five classifiers (logistic regression [LR], linear discriminant analysis [LDA], random forest [RF], support vector machine [SVM], and gradient boosting [GB]) with nested cross-validation, repeated cross-validation and a permutation test; (iii) triangulates SHAP-based explainability with univariate effect sizes, collinearity diagnostics and domain-restricted model comparisons; and (iv) examines whether discrimination persists in performance-only domain subsets that exclude age, training age, height, and body mass. We hypothesised that (i) domains beyond isokinetic strength (training history, vertical jump, Wingate power, sprint, agility) would contribute substantially to discrimination; (ii) relative isokinetic strength and power would remain significant but would not be the dominant domain once the fuller battery was modelled; and (iii) SHAP would corroborate univariate and domain-restricted evidence and reveal that raw multivariate coefficient magnitude is unreliable when deterministically derived indices (H:Q ratio, CMJ–SJ difference, SSC%) are entered alongside the raw components from which they are computed.

2. Materials and Methods

2.1. Dataset and Participants

A cross-sectional dataset of 201 male soccer players was analysed, comprising 101 amateur and 100 professional players (proportion professional = 0.498). No observations were missing for any of the 45 study variables.

Professional players were defined as players holding a professional contract with a club competing in the Turkish professional leagues (Süper Lig, 1. Lig, 2. Lig or 3. Lig) and amateur players as licensed players of clubs competing in amateur leagues. Inclusion criteria were age $\geq$ 18 years, no lower-limb injury in the preceding six months and full participation in team training. Playing position was not considered in the analyses (see Limitations).

Players were recruited by convenience sampling: professional players from clubs competing in the four Turkish professional leagues and amateur players from clubs competing in amateur leagues. Club membership was not retained in the anonymised dataset, so the number of clubs and the distribution of players across leagues cannot be reported (see Limitations). Players who met the inclusion criteria and completed the full test battery were included; incomplete test batteries were not entered into the dataset, which therefore comprises 201 players with complete data. Each observation represents a different player, and no player was tested more than once.

The between-group differences in age, training age, stature and body mass were not targeted by design. They reflect the sampling framework, in which status was defined by contract and league rather than by matching on these characteristics, and the recruitment sources (senior professional squads versus amateur club squads). Both groups were tested during the first half of the competitive season, in the same laboratory, with the same equipment and protocol and by the same research team; the assessors were not blinded to competitive level.

The study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Commission of Gazi University. All participants were informed about the procedures and gave written informed consent; the approval covered the secondary research use of the test data, which were anonymised before analysis.

2.2. Testing Procedures

Testing was conducted under standardised conditions. Players were instructed to avoid strenuous exercise for 24 h before testing and completed a standardised warm-up of low-intensity running, dynamic stretching and progressive accelerations, followed by familiarisation trials for each test. All tests were completed in two sessions separated by 48–72 h. Session 1 comprised anthropometry, squat and countermovement jumps, repeated hopping, linear sprints and the reactive-agility test, in that order. Session 2 comprised isokinetic testing followed, after at least 15 min of passive recovery, by the Wingate test, so that tests were sequenced from the least to the most fatiguing and the anaerobic power test was performed last. At least 3–5 min of recovery separated consecutive tests within a session. Stature and body mass were measured with a wall-mounted stadiometer (seca 216; seca GmbH & Co. KG, Hamburg, Germany) and a calibrated digital scale (seca 813; seca GmbH & Co. KG, Hamburg, Germany). Training age was defined as the self-reported number of years of organised, club-based soccer training.

Isokinetic strength: Concentric knee extension and flexion peak torque of both legs were measured on an isokinetic dynamometer (IsoMed 2000; D. & R. Ferstl GmbH, Hemau, Germany) at $60°\cdot s^{-1}$ and $180°\cdot s^{-1}$. Players were seated with the trunk, pelvis and thigh stabilised and the dynamometer axis aligned with the knee joint. The range of motion was set from 90° of knee flexion to 10° short of full extension (0° = full extension; 80° arc). Velocities were tested in a fixed order, from the slower ($60°\cdot s^{-1}$) to the faster ($180°\cdot s^{-1}$), and the dominant leg was tested first. At each velocity, three submaximal familiarisation repetitions were followed by five maximal reciprocal extension–flexion repetitions, with 60 s of rest between velocities and 3 min between legs; standardised verbal encouragement was given and torque was gravity-corrected. The highest peak torque at each velocity and side was retained and normalised to body mass (PT/W).

Jumps and leg stiffness: SJ and CMJ were performed with the hands on the hips on an optical jump-measurement system (OptoJump; Microgate, Bolzano, Italy) that derives jump height from flight time. For the CMJ, players started from an upright stance, performed a fast downward movement to a self-selected depth immediately followed by a maximal vertical jump, and were instructed to keep the legs extended during flight and to land at the take-off point. The SJ started from a static semi-squat (≈90° knee flexion) held for about 2 s; trials showing a visible countermovement were repeated, and the best of three trials was retained [18]. For both jumps, trials were separated by about 60 s of passive recovery. Because the variable labelled EUR in the original dataset is the absolute CMJ−SJ height difference (cm) rather than the conventional CMJ:SJ ratio [16], it is referred to throughout this manuscript as the CMJ–SJ difference; its group means equal the differences between the CMJ and SJ means (Table 1). SSC% was calculated as $100 \times (\mathrm{CMJ} - \mathrm{SJ})/\mathrm{SJ}$. Leg stiffness was derived from contact and flight times recorded with the same system (OptoJump with OptoJump Next software, version 1.15.0) in a repeated hopping test: players performed two sets of seven consecutive bilateral hops on the spot at a self-selected frequency, with the hands on the hips, and were instructed to keep the knees as extended as possible and to minimise ground contact time; the first two hops of each set were discarded and the remaining hops were averaged. Leg stiffness ($kN\cdot m^{-1}$) was calculated with the spring-mass equation of Dalleau et al. [58], $k = [M\cdot\pi\cdot(t_f + t_c)]/[t_c^2\cdot((t_f + t_c)/\pi - t_c/4)]$, where M is body mass, t_c is contact time and t_f is flight time.

Wingate anaerobic power: Peak power (PP), average power (AP), minimum power (MP), time to peak power and power decrement (PD) were obtained from a 30 s all-out Wingate cycle-ergometer test (Monark 894E Peak Bike; Monark Exercise AB, Vansbro, Sweden) with a braking load of $0.075\ kg\cdot kg^{-1}$ body mass; MP denotes the minimum power output and $PD = 100 \times (PP - MP)/PP$. Time to peak power was the time from the start of

the test to PP. After a standardised 5 min cycling warm-up including two short accelerations and 3 min of rest, players pedalled maximally from a rolling start, with the load applied at the start of the test and verbal encouragement throughout. These cycle-ergometer outcomes are reported as a separate Wingate anaerobic power domain and are not jump-derived measures. The observed values (PP 11.9–13.1 $W \cdot kg^{-1}$; AP 8.3–9.3 $W \cdot kg^{-1}$) are consistent with Wingate norms for adult soccer players [59].

Sprint and agility: Sprint times over 5, 10, 20 and 30 m were recorded with dual-beam photoelectric timing gates (Witty; Microgate, Bolzano, Italy) from a standing split-stance start with the front foot 0.5 m behind the first gate; gates were set at approximately hip height (≈1.0 m), two trials were separated by at least 3 min of passive recovery, and the best trial was retained [31]. Reactive agility was assessed with a test in which players responded to unpredictable visual stimuli [60]; total time and the left, right and mean 180° turn times were recorded by the test system (SpeedCourt; Globalspeed GmbH, Hemau, Germany). The SpeedCourt consists of a 5.25 m × 5.25 m platform with twelve contact plates positioned in a symmetric order and a large screen in front of the player that indicates the next target plate [61]. In each trial, players started from the centre of the platform and, as quickly as possible, ran to and touched the target plate indicated randomly on the screen and returned to the centre, continuing until the stimulus sequence had been completed; the sequence was predefined by the system software and identical for all players. Two trials were performed with 3 min of passive recovery, and the best total time was retained. Turn time was defined as the contact time on the target plate at which the 180° change of direction was performed, from initial foot contact to take-off, averaged separately for targets on the right and left sides; the mean turn time was the average of the two sides.

Reliability: Test–retest reliability was not assessed within the present dataset, so the measurement precision achieved under the present procedures is unknown. The tests used in this study have shown good-to-excellent test–retest reliability in published studies: isometric and isokinetic knee flexion and extension peak torque (intraclass correlation coefficient [ICC] 0.90–0.98 on the IsoMed 2000) [62], SJ and CMJ height (Cronbach's α 0.97–0.98; coefficient of variation [CoV] 2.4–4.6% across jump tests) [63], Wingate power indices (peak power ICC 0.99, CoV ≈ 1%; average power and the other traditional indices ICC ≥ 0.94, CoV ≤ 2.8%) [64], photocell sprint times [31] and soccer-specific reactive-agility tests [30]. These estimates were obtained in other samples and laboratories; they provide external context for the expected precision of the instruments but do not establish the reliability of the present measurements, which is acknowledged as a limitation when between-group differences are translated into individual benchmarks.

2.3. Variables

The target variable was competitive level (amateur = 0; professional = 1). Forty-five candidate physical performance variables were available across six domains (complete data dictionary in Appendix A, Table A1): (1) anthropometry/training history: age, training age, height, body mass; (2) isokinetic strength, absolute: concentric knee flexion and extension peak torque (PT, Nm) at $60° \cdot s^{-1}$ and $180° \cdot s^{-1}$, right and left; (3) isokinetic strength, relative: the same eight values normalised to body mass (PT/W, $Nm \cdot kg^{-1}$); (4) H:Q ratio: flexion:extension torque ratio (%) at both velocities, bilateral; (5) jump performance and Wingate anaerobic power, two distinct constructs: (5a) vertical jump and elastic energy—CMJ and SJ height, CMJ–SJ difference (cm), SSC% and leg stiffness—and (5b) Wingate cycle-ergometer power—peak, average and minimum power in absolute and relative units (PP, AP, MP; W and $W \cdot kg^{-1}$), time to peak power and power decrement (PD%); (6) agility and sprint: reactive agility time, mean/left/right 180° turn time, and sprint time over 5, 10, 20 and 30 m. For the domain-restricted analyses, isokinetic strength was represented by

PT/W only, H:Q ratios were excluded (Section 2.4), and agility and sprint were combined into a single sprint-and-agility domain. The jump and Wingate features were analysed as one eight-feature subset, referred to as the jump-plus-Wingate-power subset; because it combines two constructs, it is not interpreted as a jump-derived measure.

2.4. Feature Curation and Multicollinearity Screening

Feature curation used two distinct types of rules, which are reported separately (Table A1). (1) Exclusions were fixed by the mathematical definition of the variables, which do not depend on the sample. The H:Q ratio equals 100 × (PT/W flexion)/(PT/W extension) at the same velocity and side, and SSC% equals 100 × (CMJ − SJ)/SJ; both are deterministic but nonlinear (quotient) functions of retained variables. The CMJ–SJ difference equals CMJ − SJ and is therefore an exact linear combination of two retained features. As a data-integrity check, the stored H:Q and SSC% values were correlated with the same quantities recomputed from their components ($r = 0.9999$ and $r = 0.999999$, respectively). Absolute isokinetic torque (Nm) and absolute power outputs (W) were excluded in favour of their body-mass-normalised equivalents ($Nm \cdot kg^{-1}$, $W \cdot kg^{-1}$), because body mass was entered as a separate feature. (2) Exclusions were based on sample-estimated correlations. The individual left- and right-turn times were replaced by the mean 180° turn time ($r > 0.9999$ with each side), and the 10 m and 20 m splits were excluded because of their high correlations with the retained 5 m (acceleration) and 30 m (maximal-velocity approach) splits ($r = 0.78$–0.93).

These rules address three different phenomena. Algebraic dependence means that a variable is an exact function of other variables (linear for the CMJ–SJ difference; nonlinear for H:Q and SSC%). Linear multicollinearity means that predictors are strongly but not perfectly correlated in the sample (e.g., CMJ and SJ height, adjacent sprint splits). Suppression refers to a predictor whose coefficient becomes large, or changes sign, mainly because it removes variance shared with other predictors [48,54]. The exclusions were model-specific rather than a general prohibition of derived variables. In the linear models, entering the CMJ–SJ difference together with CMJ and SJ makes the design matrix rank-deficient, so the three coefficients are not separately identifiable. Over the narrow observed range, the nonlinear indices (H:Q, SSC%) are close to linear functions of their components, so their coefficients would mainly express contrasts between components rather than an independent construct. Tree-based models tolerate derived variables, but retaining both a ratio and its components would split importance across redundant representations and complicate comparison of attributions across models. A derived index may legitimately be modelled instead of, rather than in addition to, its components when the derived construct is the scientific target. On this basis, Wingate power decrement was retained although it is computed from PP and MP: fatigue resistance was itself a construct of interest, and as a nonlinear quotient it showed low linear dependence on the retained Wingate variables (VIF = 1.7; Table 2), so its coefficient remained identifiable.

The curated 24-feature set (4 anthropometric/training-history, 8 isokinetic PT/W, 3 vertical-jump/leg-stiffness, 5 Wingate, 2 agility and 2 sprint features; Table 2) retained substantial linear multicollinearity between CMJ and SJ height (VIF = 21.4 and 20.1); all other features had VIF < 9. Both were retained because they measure different jump conditions. Their high VIF means that their individual coefficients and attributions are unstable and that importance may be shared or reallocated between them, so they are interpreted jointly (Section 3.4). The group label was not used in any curation decision. However, the correlation-based screening in rule (2) and the VIF diagnostics used the full sample ($n = 201$), so information from observations later held out in cross-validation or in the test set could have influenced which of the correlated sprint and turn-time variables were

retained. Because this screening was label-free and affected only the choice among highly correlated sprint-split and turn-time variables, its influence on the cross-validated estimates is expected to be small; it was not repeated within training folds, which is acknowledged as a limitation (Section 4).

2.5. Statistical Analysis

For each of the 45 variables, amateur and professional groups were compared with Welch's *t*-test (unequal variances) and, as a non-parametric sensitivity check, the Mann–Whitney U test. Because variances differed markedly between groups for several timing variables (standard deviation [SD] ratio up to 3.6:1, amateurs more variable), pooled-SD effect sizes for these variables are interpreted alongside the Welch test, which does not assume equal variances. Cohen's d (pooled SD) with 95% confidence intervals (CIs; Hedges and Olkin approximation) quantified effect size [65]. To control the expected proportion of false positives among the significant results across the 45 comparisons (the false-discovery rate, not the family-wise error rate), *p*-values were adjusted with the Benjamini–Hochberg false-discovery-rate (FDR) procedure [66]. Univariate discrimination was quantified with the area under the receiver operating characteristic (ROC) curve (AUC), using the variable's raw values as the classifier score and reporting the direction-agnostic magnitude (max[AUC, 1−AUC]) alongside the professional-favouring direction. Pairwise Pearson correlations and variance inflation factors (VIFs) were computed for the curated 24-feature set to diagnose multicollinearity [48,50–53]. With $n = 101$ and 100, the study had 80% power (two-sided $\alpha = 0.05$) to detect differences of $d \geq 0.40$ ($d \geq 0.59$ at a Bonferroni-equivalent α of 0.05/45); smaller differences could be neither reliably detected nor excluded. Overall multivariate separation was expressed as a model-implied Mahalanobis distance, $D = \sqrt{2} \cdot \Phi^{-1}(\mathrm{AUC})$, which links the AUC of the linear discriminant to D when two multivariate normal groups share a common covariance matrix [67,68].

2.6. Model Development

The 24 curated features were standardised (z-transformation fitted on the training fold only, to avoid leakage) and split into training (70%) and test (30%) sets with stratification (random_state = 42; training $n = 140$, test $n = 61$). The complete analytical workflow, including the observations used for each analysis, is shown in Appendix A (Figure A1). Five classifiers were fitted: logistic regression (LR), linear discriminant analysis (LDA), random forest (RF) [69], support vector machine with a radial basis function kernel (SVM), and gradient boosting (GB) [70]. Hyperparameters (regularisation strength for LR/SVM, tree depth and leaf size for RF/GB; LR used an L2 (ridge) penalty, the scikit-learn default, with C tuned over a logarithmic grid; the grids and selected values are provided with the analysis code) were tuned by grid search within a 5-fold stratified inner cross-validation (CV) loop, nested inside a 5-fold stratified outer loop, optimising ROC-AUC. This nested-CV design was chosen to obtain a near-unbiased performance estimate given the modest sample size [71–73]. Nested cross-validation was performed within the 140-player development set. The final hyperparameters of each classifier were selected by repeating the inner 5-fold grid search on the complete development set; the outer loop was used only to estimate performance. The 61 test players were not used for model fitting, hyperparameter tuning or threshold selection. They were, however, included in the label-free feature curation (Section 2.4) and in the full-sample supplementary analyses (repeated cross-validation, permutation test, composite index, domain-restricted comparisons and PCA; Figure A1), which are therefore not independent of the test set. SVM class probabilities, required for ROC-AUC and the Brier score, were obtained by Platt scaling with internal five-fold cross-validation as implemented in scikit-learn (SVC, probability = True) [74]. Final models were

refit on the full training set with the best hyperparameters and evaluated once on the held-out test set (accuracy, precision, recall, F1, ROC-AUC, precision–recall AUC, Brier score).

Because the two classes were almost completely separable, unpenalised maximum-likelihood logistic coefficients would not be finite [75]; all reported LR coefficients and odds ratios are therefore penalised (shrunken) estimates and are interpreted as relative per-SD associations rather than conventional effect estimates. With 24 candidate predictors and 70 professional players in the development set (≈2.9 events per predictor), the sample was below the minimum suggested by Riley et al.'s criteria (≈470 and ≈385 participants for anticipated C-statistics of 0.90 and 0.99, respectively, assuming 24 parameters, an outcome proportion of 0.50 and a target shrinkage of 0.90) [76]. This shortfall reinforces the need for penalisation, internal validation and cautious interpretation of individual coefficients. Because analytical confidence intervals for the AUC (e.g., DeLong's method [77]) become unreliable or degenerate as the AUC approaches 1.0, AUC uncertainty is summarised by the 2.5th–97.5th percentile range across the 100 repeated cross-validation folds; because these folds share training observations and are not independent replications, this range describes the resampling variability of fold-level estimates and is not a confidence interval for population performance [78]; test-set accuracy is reported with Wilson 95% confidence intervals [79]. Reporting follows the TRIPOD+AI (Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis, updated for artificial intelligence) statement [80] (completed checklist in the Supplementary Table S1).

2.7. Robustness Checks

Several additional analyses examined the internal stability of the near-ceiling discrimination. They can reveal instability across data splits or gross pipeline errors but cannot exclude systematic differences between the recruited groups (e.g., club, recruitment or measurement differences). First, 20 repetitions of stratified five-fold CV on the full sample (n = 201; 100 held-out folds per model), with hyperparameters fixed at values selected previously from the same data, quantified the spread of AUC across resampling for LR, LDA and RF. Because the hyperparameters were not re-tuned within each repetition, this analysis is an internal stability assessment of a fixed model configuration, not a repeated nested evaluation of the full model-development procedure. SVM and GB were not included in this check, but all five classifiers reached comparable nested-CV AUCs (0.994–0.999; Section 3.3). Second, a permutation test shuffled the class labels 1000 times and refit the LR pipeline under 5-fold cross-validation each time (with the regularisation parameter fixed, i.e., without re-tuning within each permutation), yielding a null AUC distribution against which the true, unpermuted score was ranked. The scaler was refitted within every training fold as part of the pipeline, the observed score was the 5-fold cross-validated AUC on the full sample, and p was computed as $(b + 1)/(m + 1)$, where b is the number of permuted AUCs $\geq$ the observed AUC and m = 1000, as implemented in scikit-learn [74,81,82]. Third, an unsupervised composite index was built by z-scoring all 24 features (sign-flipping those where a lower value is better: time to peak-power, reactive agility, mean-turn time, 5 m and 30 m sprint) and averaging them with equal, unfitted weights; its AUC against the professional/amateur label provides a fitting-free sanity check on the separability of the two groups. This index is not entirely assumption-free, because the direction of the four anthropometric/training-history variables was set from the expected professional advantage.

2.8. Feature Importance and Visualisation

Feature importance was interpreted three ways: standardised LR coefficients (β) with odds ratios ($OR = e^{\beta}$) and LDA discriminant coefficients; SHapley Additive exPlanations (SHAP) [55] computed with a tree explainer for the better performing of RF/GB (selected

by nested-CV AUC; TreeSHAP with tree-path-dependent feature perturbation, the default when no background dataset is supplied, computed for the training-set observations); SHAP values are on the probability scale for the professional class (class 1) and features were ranked by global mean absolute SHAP value; and domain-restricted model comparisons comparing the 5-fold cross-validated performance of the anthropometric-only, isokinetic-PT/W-only, jump-plus-Wingate-power-only, sprint-and-agility-only, CMJ+SJ-only, and PP[W/kg]-only feature subsets against the full 24-feature model, for both LR and RF (single 5-fold cross-validation on the full sample, with hyperparameters tuned by an inner grid search within each training fold; subsets of equal size were also compared directly). These standalone comparisons show how well each restricted feature set discriminates on its own; they do not quantify the incremental contribution of each domain to the full model. Each value is a point estimate from a single five-fold CV, so differences of 0.02 or less between subsets were not interpreted. Principal component analysis (PCA; two components, fitted on the standardised full-sample data) provided a two-dimensional visualisation of group separation. Because SHAP attributions can be distorted when features are correlated [56,57], they were interpreted alongside univariate effect sizes and penalised LR coefficients rather than in isolation. Because correlated predictors share predictive information, SHAP may distribute attribution among them differently depending on model structure; the rankings therefore describe contributions to classification within this dataset, not causal or independent physiological importance.

2.9. Software

All analyses were conducted in Python 3.14 (Python Software Foundation, Wilmington, DE, USA) using pandas 3.0, scikit-learn 1.9 [74], SciPy 1.18, statsmodels 0.14 and SHAP 0.52 [55]; figures used matplotlib 3.11 and seaborn 0.13. All random seeds were fixed (random_state = 42) for reproducibility. The analysis code is available from the corresponding author on request (see Data Availability Statement).

3. Results

3.1. Descriptive Statistics and Univariate Comparisons

Table 1 and Figure 1 summarise group means, standard deviations, and univariate comparisons for all 45 variables, organised by domain; Figure 2 displays the corresponding effect sizes as a forest plot.

After FDR correction, every variable was significantly different between groups (all p_FDR < 0.001) except the four H:Q ratios, the CMJ–SJ difference and SSC% (all p_FDR ≥ 0.28), whose effect sizes were close to zero (d = −0.10 to 0.16) and whose univariate AUC was at chance level (0.51–0.57). Their 95% CIs (upper limits 0.18–0.44) do not exclude small differences, so these are absence-of-evidence rather than evidence-of-absence findings (Section 2.5). Among the significant variables, effect sizes ranged from d = 0.49 (power decrement) to d = 2.10 (training age). Several non-isokinetic variables showed the largest effects in the battery: training age (d = 2.10, AUC = 0.92), Wingate average power (d = 2.04, AUC = 0.93), CMJ height (d = 2.02, AUC = 0.93), Wingate peak power in absolute watts (d = 1.89, AUC = 0.91), reactive agility (d = −1.69, AUC = 0.88) and 30 m sprint time (d = −1.52, AUC = 0.88). By comparison, the eight relative isokinetic peak-torque variables (PT/W) showed smaller, though still significant, effects (d = 0.70–0.97, AUC = 0.69–0.76). Absolute isokinetic peak torque, however, showed effects (d = 1.41–1.86) comparable to those of sprint and agility, indicating that much of the professional advantage in torque is accounted for by body mass (d = 1.59).

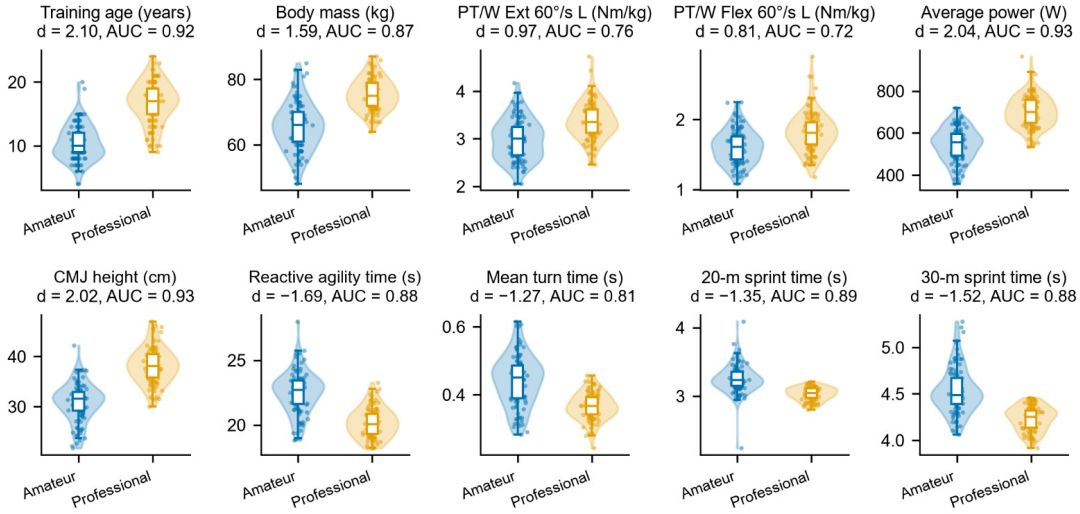


Figure 1. Group distributions of the two strongest discriminators per performance domain (violin + box + individual data points). d = Cohen's d; AUC = area under the receiver operating characteristic curve; CMJ = countermovement jump; PT/W = peak torque normalised to body mass; Ext = extension; Flex = flexion; L = left.



Figure 2. Forest plot of Cohen's d (95% CI) for all 45 variables, colour-coded by domain. Vertical-jump and Wingate variables are shown as separate domains.

Table 1. Descriptive statistics and univariate comparisons for all 45 physical performance variables, organised by domain (Benjamini–Hochberg FDR-corrected p; Sig.: *** $p < 0.001$, n.s. = not significant).

Domain	Variable	Amateur (M ± SD)	Professional (M ± SD)	Cohen's d (95% CI)	p (FDR)	Sig.
Anthropometric	Training age (years)	10.45 ± 2.79	16.80 ± 3.25	2.10 (1.75, 2.44)	<0.001	***
	Body mass (kg)	65.65 ± 7.28	75.45 ± 4.81	1.59 (1.27, 1.90)	<0.001	***
	Height (cm)	172.15 ± 8.80	181.83 ± 6.77	1.23 (0.93, 1.53)	<0.001	***
	Age (years)	21.93 ± 2.18	24.75 ± 2.78	1.13 (0.83, 1.43)	<0.001	***
Isokinetic strength (absolute, Nm)	PT Ext 60°/s L (Nm)	196.16 ± 37.10	253.97 ± 23.70	1.86 (1.52, 2.19)	<0.001	***
	PT Ext 60°/s R (Nm)	202.92 ± 37.75	255.39 ± 25.46	1.63 (1.31, 1.95)	<0.001	***
	PT Ext 180°/s L (Nm)	167.46 ± 29.89	212.37 ± 25.75	1.61 (1.29, 1.93)	<0.001	***
	PT Flex 60°/s L (Nm)	105.77 ± 20.68	136.69 ± 18.17	1.59 (1.27, 1.91)	<0.001	***
	PT Ext 180°/s R (Nm)	167.53 ± 32.89	213.94 ± 26.88	1.54 (1.23, 1.86)	<0.001	***
	PT Flex 180°/s L (Nm)	96.69 ± 19.66	124.59 ± 17.66	1.49 (1.18, 1.81)	<0.001	***
	PT Flex 60°/s R (Nm)	107.05 ± 23.68	137.65 ± 17.77	1.46 (1.15, 1.77)	<0.001	***
	PT Flex 180°/s R (Nm)	95.81 ± 23.84	125.46 ± 17.62	1.41 (1.10, 1.72)	<0.001	***
Isokinetic strength (relative, Nm/kg)	PT/W Ext 60°/s L (Nm/kg)	2.98 ± 0.43	3.38 ± 0.38	0.97 (0.68, 1.27)	<0.001	***
	PT/W Flex 60°/s L (Nm/kg)	1.61 ± 0.24	1.82 ± 0.28	0.81 (0.52, 1.09)	<0.001	***
	PT/W Ext 60°/s R (Nm/kg)	3.08 ± 0.41	3.40 ± 0.40	0.77 (0.49, 1.06)	<0.001	***
	PT/W Ext 180°/s L (Nm/kg)	2.55 ± 0.37	2.82 ± 0.37	0.74 (0.45, 1.02)	<0.001	***
	PT/W Ext 180°/s R (Nm/kg)	2.55 ± 0.41	2.84 ± 0.38	0.73 (0.45, 1.02)	<0.001	***
	PT/W Flex 180°/s R (Nm/kg)	1.46 ± 0.31	1.67 ± 0.27	0.74 (0.46, 1.03)	<0.001	***
	PT/W Flex 60°/s R (Nm/kg)	1.63 ± 0.28	1.83 ± 0.28	0.75 (0.46, 1.03)	<0.001	***
	PT/W Flex 180°/s L (Nm/kg)	1.47 ± 0.26	1.66 ± 0.28	0.70 (0.41, 0.98)	<0.001	***
H:Q ratio (%)	H:Q ratio 180°/s R (%)	57.68 ± 11.67	59.12 ± 8.31	0.14 (−0.13, 0.42)	0.342	n.s.
	H:Q ratio 60°/s R (%)	52.90 ± 7.51	53.93 ± 4.78	0.16 (−0.11, 0.44)	0.276	n.s.
	H:Q ratio 180°/s L (%)	58.16 ± 9.49	59.12 ± 8.38	0.11 (−0.17, 0.38)	0.467	n.s.
	H:Q ratio 60°/s L (%)	54.45 ± 8.14	53.81 ± 4.88	−0.10 (−0.37, 0.18)	0.510	n.s.
Jump performance and Wingate anaerobic power	Wingate average power (W)	546.17 ± 78.07	701.96 ± 74.97	2.04 (1.69, 2.38)	<0.001	***
	Countermovement jump height (cm)	30.90 ± 3.71	38.23 ± 3.55	2.02 (1.68, 2.36)	<0.001	***
	Squat jump height (cm)	30.29 ± 3.79	37.45 ± 3.71	1.91 (1.58, 2.24)	<0.001	***
	Wingate peak power (W)	782.50 ± 120.83	988.86 ± 95.72	1.89 (1.56, 2.23)	<0.001	***
	Wingate minimum power (W)	316.83 ± 56.49	403.31 ± 47.35	1.66 (1.34, 1.98)	<0.001	***
	Wingate average power (W/kg)	8.30 ± 0.68	9.30 ± 0.78	1.37 (1.06, 1.67)	<0.001	***
	Wingate peak power (W/kg)	11.90 ± 1.22	13.11 ± 0.95	1.10 (0.80, 1.40)	<0.001	***
	Wingate time to peak power (ms)	1561.96 ± 518.64	1189.60 ± 145.99	−0.98 (−1.27, −0.68)	<0.001	***
	Wingate minimum power (W/kg)	4.82 ± 0.67	5.35 ± 0.52	0.88 (0.59, 1.17)	<0.001	***
	Leg stiffness (kN/m)	21.66 ± 4.97	24.52 ± 4.20	0.62 (0.34, 0.90)	<0.001	***
	Wingate Power decrement (%)	58.80 ± 5.77	56.00 ± 5.73	−0.49 (−0.77, −0.21)	<0.001	***
	CMJ–SJ difference (cm) [a]	0.61 ± 1.45	0.79 ± 1.08	0.14 (−0.14, 0.41)	0.356	n.s.
	Stretch-shortening cycle contribution (%)	2.22 ± 5.28	2.21 ± 3.00	−0.00 (−0.28, 0.27)	0.982	n.s.
Agility	Reactive agility time (s)	22.51 ± 1.60	20.17 ± 1.14	−1.69 (−2.01, −1.37)	<0.001	***
	Mean turn time (s)	0.44 ± 0.07	0.36 ± 0.04	−1.27 (−1.57, −0.97)	<0.001	***
	Mean right-turn time (s)	0.43 ± 0.08	0.36 ± 0.05	−1.09 (−1.38, −0.79)	<0.001	***
	Mean left-turn time (s)	0.45 ± 0.09	0.37 ± 0.06	−1.02 (−1.32, −0.73)	<0.001	***
Sprint	20 m sprint time (s)	3.25 ± 0.21	3.03 ± 0.10	−1.35 (−1.66, −1.05)	<0.001	***
	30 m sprint time (s)	4.53 ± 0.25	4.23 ± 0.13	−1.52 (−1.83, −1.20)	<0.001	***
	10 m sprint time (s)	1.94 ± 0.15	1.81 ± 0.06	−1.17 (−1.47, −0.87)	<0.001	***
	5 m sprint time (s)	1.17 ± 0.13	1.09 ± 0.04	−0.85 (−1.14, −0.56)	<0.001	***

Note. M = mean; SD = standard deviation; CI = confidence interval; FDR = Benjamini–Hochberg false-discovery-rate corrected p-value; PT = peak torque; Ext = extension; Flex = flexion; L = left; R = right; PT/W = peak torque normalised to body mass; H:Q = hamstring:quadriceps ratio; CMJ = countermovement jump; SJ = squat jump; Wingate = 30 s Wingate cycle-ergometer test; SSC% = stretch-shortening-cycle contribution. Sig.: *** $p < 0.001$, n.s. = not significant. Effect sizes were calculated from unrounded data; means and SDs of times are rounded to two decimals. [a] CMJ height minus SJ height; labelled "eccentric utilisation ratio" in the original dataset, but not the conventional CMJ:SJ ratio [16].

3.2. Multicollinearity Diagnostics

Table 2 reports the curated 24-feature set and its variance inflation factors. Once absolute PT, H:Q ratio, CMJ–SJ difference, SSC%, the individual left/right-turn times and the two intermediate sprint splits were removed for the reasons detailed in Methods, only the CMJ–SJ pair retained a VIF above 10 (21.4 and 20.1); all other features had VIF < 9 (Figure 3). The CMJ–SJ difference and SSC% were excluded on algebraic grounds: the CMJ–SJ difference is an exact linear combination of CMJ and SJ height, and the stored SSC% values matched 100 × (CMJ − SJ)/SJ to six decimal places (r = 0.999999). Both were statistically indistinguishable between groups on their own (Table 1, n.s.) even though they depend almost perfectly on two variables (CMJ and SJ height) that individually show some of the largest effect sizes in the battery (d = 2.02 and 1.91). Entered alongside CMJ and SJ, such variables would make the linear models rank-deficient (CMJ–SJ difference) or nearly so (SSC%), producing non-identifiable or unstable coefficients; both were excluded and retained only for univariate reporting. The H:Q ratio is a deterministic but nonlinear (quotient) function of the PT/W flexion and extension values already included; it was excluded from the multivariable models because, with both components present, its coefficient would mainly express a contrast between them rather than an independent construct (Section 2.4). Because CMJ and SJ height remain strongly collinear (VIF > 20), their individual coefficients and SHAP attributions are interpreted jointly.

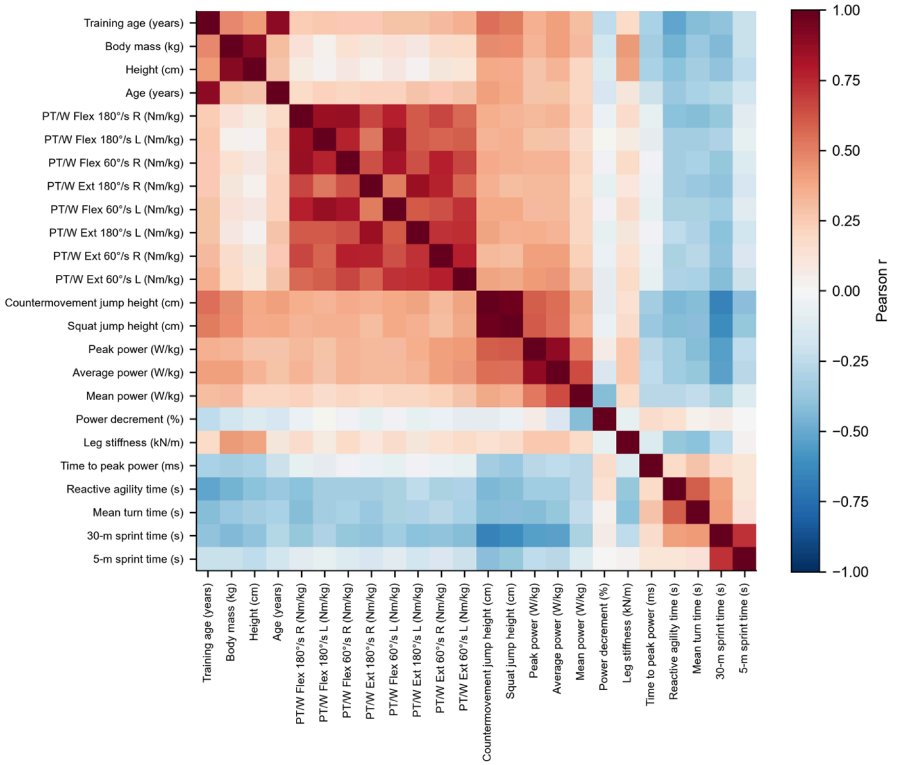


Figure 3. Correlation matrix of the 24 curated model features.

Table 2. Curated 24-feature model set and variance inflation factors (VIFs).

Domain	Feature	VIF	Interpretation
Vertical jump/leg stiffness	Countermovement jump height (cm)	21.4	Elevated (CMJ–SJ pair)
Vertical jump/leg stiffness	Squat jump height (cm)	20.1	Elevated (CMJ–SJ pair)
Isokinetic strength (relative, Nm/kg)	PT/W Flex 180°/s R (Nm/kg)	8.9	Moderate
Anthropometric	Training age (years)	8.4	Moderate

Table 2. *Cont.*

Domain	Feature	VIF	Interpretation
Anthropometric	Body mass (kg)	8.0	Moderate
Isokinetic strength (relative, Nm/kg)	PT/W Flex 180°/s L (Nm/kg)	7.5	Moderate
Isokinetic strength (relative, Nm/kg)	PT/W Flex 60°/s R (Nm/kg)	7.3	Moderate
Isokinetic strength (relative, Nm/kg)	PT/W Ext 180°/s R (Nm/kg)	7.1	Moderate
Wingate anaerobic power	Peak power (W/kg)	7.0	Moderate
Wingate anaerobic power	Average power (W/kg)	6.9	Moderate
Isokinetic strength (relative, Nm/kg)	PT/W Flex 60°/s L (Nm/kg)	6.8	Moderate
Anthropometric	Height (cm)	6.7	Moderate
Isokinetic strength (relative, Nm/kg)	PT/W Ext 180°/s L (Nm/kg)	6.7	Moderate
Isokinetic strength (relative, Nm/kg)	PT/W Ext 60°/s R (Nm/kg)	5.8	Moderate
Anthropometric	Age (years)	5.8	Moderate
Isokinetic strength (relative, Nm/kg)	PT/W Ext 60°/s L (Nm/kg)	4.8	Low
Sprint	30 m sprint time (s)	4.6	Low
Sprint	5 m sprint time (s)	2.9	Low
Wingate anaerobic power	Minimum power (W/kg)	2.3	Low
Agility	Reactive agility time (s)	2.2	Low
Agility	Mean turn time (s)	2.1	Low
Wingate anaerobic power	Power decrement (%)	1.7	Low
Vertical jump/leg stiffness	Leg stiffness (kN/m)	1.6	Low
Wingate anaerobic power	Time to peak power (ms)	1.4	Low

Note. VIF = variance inflation factor; PT/W = peak torque normalised to body mass. VIF > 10 indicates elevated multicollinearity risk; the CMJ–SJ pair (VIF = 21.4, 20.1) reflects strong linear multicollinearity; their individual coefficients and attributions are therefore unstable and are interpreted jointly (Section 2.4).

3.3. Classification Performance

Table 3 and Figure 4 show test-set performance for all five classifiers on the curated 24-feature model. All five achieved near-ceiling discrimination (ROC-AUC 0.998–1.000; accuracy 0.967–1.000; Wilson 95% CI lower limits 0.89–0.94), with gradient boosting reaching a perfect test-set separation (AUC = 1.000, Brier score = 0.002) and the remaining four models within AUC = 0.998–0.999. Confusion matrices (Figure 5) show 0–2 misclassifications out of 61 test-set players for every model. With at most two misclassifications per model and overlapping confidence intervals, the test set could not distinguish between the classifiers, which are therefore not ranked. Nested cross-validation (Table 4) indicated that this was not attributable to the particular 70/30 split: AUC ranged from 0.994 (GB) to 0.999 (LR), with accuracy 0.943–0.979. Twenty repetitions of stratified five-fold CV with fixed hyperparameters (100 non-independent held-out folds; Figure 6A), an internal stability assessment, showed the same near-ceiling performance with very narrow spread for LR, LDA and RF (AUC 0.996–0.999, SD $\leq$ 0.005, fold-level 2.5th–97.5th percentile range 0.983–1.000 across all three models). All five classifiers reached comparable nested-CV AUC (0.994–0.999) with small SDs across outer folds (0.002–0.012; Table 4), indicating that discrimination was insensitive to the particular model configuration.

Two further checks were consistent with this near-ceiling discrimination (Figure 6B), although neither can exclude systematic recruitment, club or measurement differences between the groups. A label-permutation test (1000 shuffles) produced a null AUC distribution centred at 0.48 (SD = 0.06), against which the true, unpermuted LR performance (AUC = 0.999) was an extreme outlier ($p = 0.001$, the smallest attainable p-value at this permutation count). An unsupervised composite index (an unweighted average of the 24 standardised, direction-corrected features, with no fitting to the group label) reached AUC = 0.969, indicating that the separability does not depend on a specific classifier or its tuning. Under multivariate normality, the cross-validated LDA AUC (0.998–0.999; Table 4)

corresponds to a model-implied Mahalanobis distance of D ≈ 4.1–4.4 [67,68], roughly twice the largest univariate effect (training age, d = 2.10).

Table 3. Test-set classification performance ($n = 61$) across five classifiers.

Model	Accuracy	Precision	Recall	F1	ROC-AUC	PR-AUC	Brier
LR	0.984	1.000	0.967	0.983	0.999	0.999	0.027
LDA	0.984	0.968	1.000	0.984	0.998	0.998	0.016
RF	0.967	0.967	0.967	0.967	0.999	0.999	0.033
SVM	0.984	1.000	0.967	0.983	0.999	0.999	0.016
GB	1.000	1.000	1.000	1.000	1.000	1.000	0.002

Note. LR = logistic regression; LDA = linear discriminant analysis; RF = random forest; SVM = support vector machine; GB = gradient boosting; ROC-AUC = area under the receiver operating characteristic curve; PR-AUC = area under the precision–recall curve; Brier = Brier score (lower is better); $n = 61$ test-set players. Wilson 95% CIs for accuracy: LR, LDA and SVM 0.913–0.997; RF 0.888–0.991; GB 0.941–1.000. Analytical AUC confidence intervals are not reported because they become unreliable or degenerate as the AUC approaches 1.0; see the repeated-CV range in Table 4.

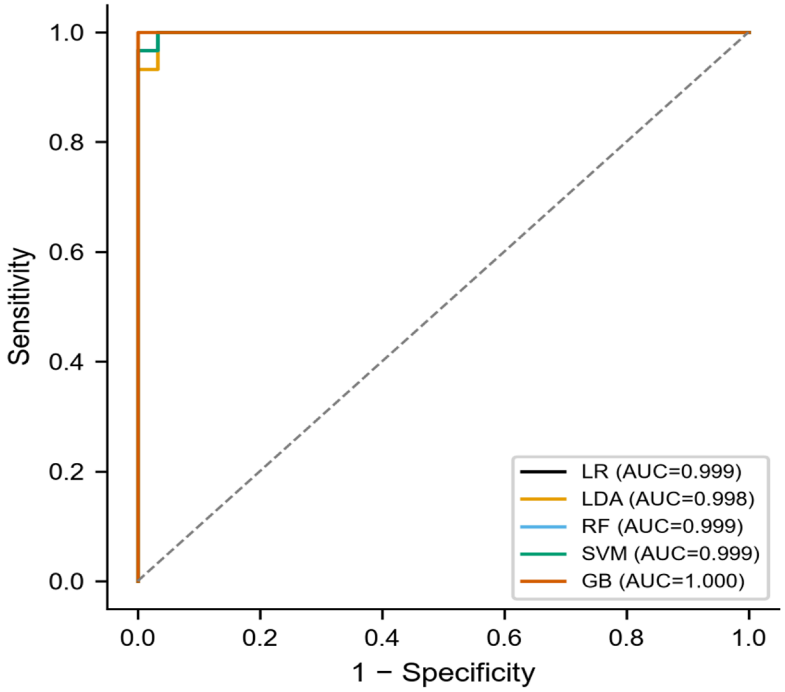


Figure 4. Test-set ROC curves across five classifiers.

Table 4. Nested (5-fold) and repeated (20 × stratified 5-fold, fixed hyperparameters) cross-validated performance.

Model	Nested CV Accuracy	Nested CV AUC	Repeated CV AUC (20 × 5-Fold)	Repeated CV AUC Fold-Level 2.5th–97.5th Percentile Range
LR	0.957 ± 0.027	0.999 ± 0.002	0.998 ± 0.003	0.990–1.000
LDA	0.943 ± 0.048	0.998 ± 0.004	0.999 ± 0.002	0.992–1.000
RF	0.971 ± 0.014	0.998 ± 0.004	0.996 ± 0.005	0.983–1.000
SVM	0.964 ± 0.023	0.996 ± 0.006	—	—
GB	0.979 ± 0.017	0.994 ± 0.012	—	—

Note. CV = cross-validation; AUC = area under the receiver operating characteristic curve. Nested CV: 5-fold outer × 5-fold inner (grid search), mean ± SD across outer folds. Repeated CV: 20 repetitions of stratified 5-fold CV with hyperparameters fixed at previously selected values (100 held-out folds), mean ± SD and fold-level 2.5th–97.5th percentile range (a descriptive measure of resampling variability, not a confidence interval); SVM and GB were not included in the repeated-CV robustness check.

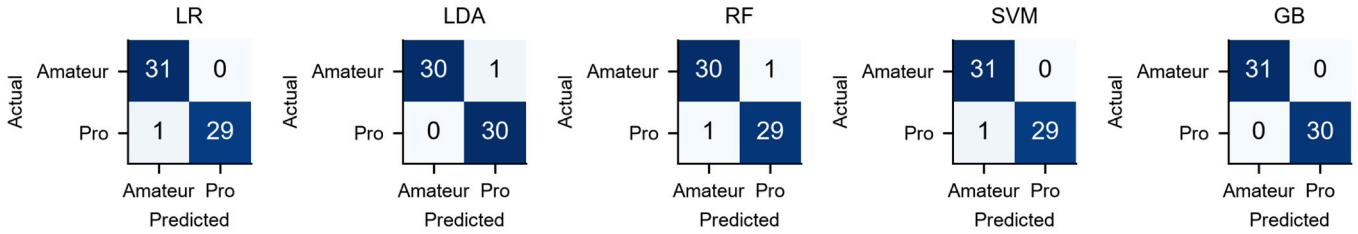


Figure 5. Test-set confusion matrices ($n = 61$).

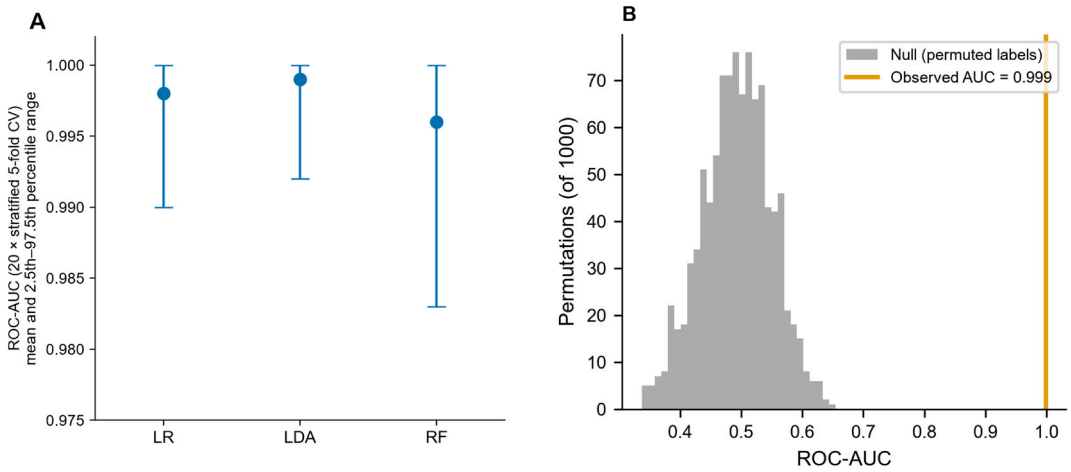


Figure 6. Robustness checks: (**A**) repeated cross-validation stability (20 × stratified 5-fold CV, fixed hyperparameters; mean and fold-level 2.5th–97.5th percentile range for LR, LDA and RF); (**B**) label-permutation test null distribution.

3.4. Feature Importance: Coefficients Versus SHAP

Table 5 and Figure 7 present feature importance from standardised LR/LDA coefficients and from SHAP (computed on the random forest, the tree model with the higher nested-CV AUC; probability scale for the professional class, ranked by mean |SHAP|). Reactive agility time, time to peak power, training age, CMJ height, squat-jump height and body mass carried the largest LR coefficients ($|\beta| = 0.40$–0.53). SHAP produced a broadly consistent, though not identical, ranking (Spearman ρ between |LR β| and mean |SHAP| = 0.82 across the 15 features in Table 5; ρ between univariate |d| and mean |SHAP| = 0.93): training age, CMJ height, squat-jump height, 30 m sprint time and reactive agility were the five highest-ranked features by mean absolute SHAP value, together accounting for a majority of total attributed importance. No feature in the curated 24-variable model showed a large multivariate coefficient alongside null univariate significance: the coefficient-based and SHAP-based rankings agreed with the univariate effect-size ranking (Table 1) after the deterministically derived variables (H:Q ratio, CMJ–SJ difference, SSC%) had been removed; this agreement is consistent with, but does not prove, the absence of suppression. The unpenalised LDA coefficients were, however, less stable: their ranking agreed only weakly with SHAP ($\rho = 0.30$), and three features (height, power decrement and PT/W flexion $60°{\cdot}s^{-1}$ left) received LDA coefficients opposite in sign to their univariate effects, a signature of residual collinearity (e.g., height with body mass; bilateral flexion torque) rather than of genuine opposite effects. Interpretation therefore rests on the penalised LR coefficients, SHAP values and univariate effect sizes, which agreed closely; the LDA coefficients are shown for completeness only. Agreement among these approaches shows that the same variables carry most of the classification signal in this dataset; it does not establish that the

highest-ranked variables are causal or independent determinants of professional status. Because CMJ and SJ height are strongly collinear, the split of attribution between them is model-dependent, and the two are best read as a single jump-height signal.

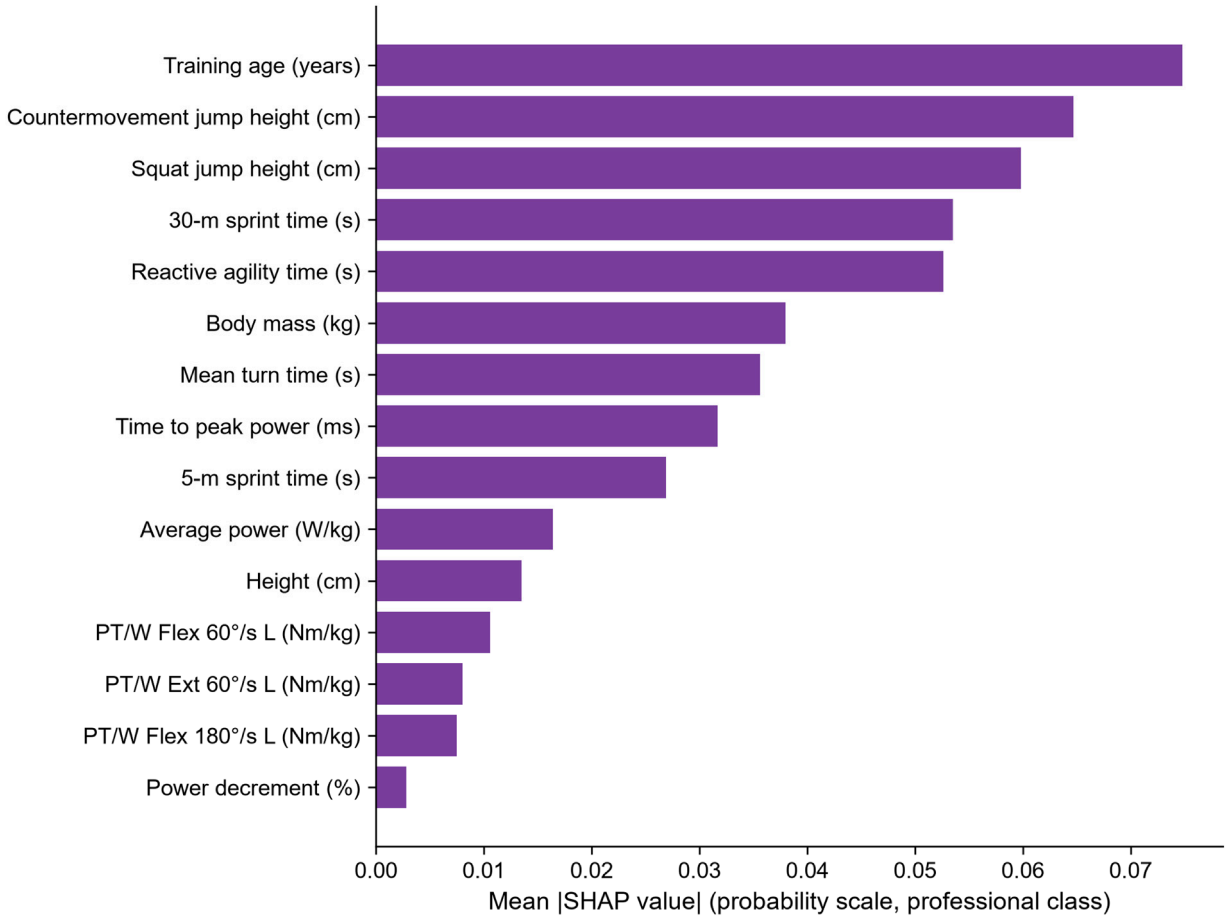


Figure 7. SHAP explainability for the random forest model: global mean |SHAP| ranking of the 15 features listed in Table 5 (probability scale, professional class; training-set observations). SHAP values quantify each feature's contribution to the model's output and should be read as model attribution, not as evidence of a causal effect.

Table 5. Standardised LR/LDA coefficients and SHAP importance for the 15 highest-ranked features (full 24-feature ranking available from the corresponding author on request).

Feature	LR β	Odds Ratio	LDA Coef.	Mean |SHAP|
Reactive agility time (s)	−0.528	0.59	−3.04	0.0526
Time to peak power (ms)	−0.527	0.59	−1.96	0.0317
Training age (years)	0.501	1.65	6.92	0.0748
Countermovement jump height (cm)	0.425	1.53	2.75	0.0647
Squat jump height (cm)	0.419	1.52	0.19	0.0598
Body mass (kg)	0.405	1.50	2.15	0.0380
30 m sprint time (s)	−0.358	0.70	−0.53	0.0535
5 m sprint time (s)	−0.346	0.71	−2.08	0.0269
Mean turn time (s)	−0.269	0.76	−0.29	0.0356
Height (cm)	0.232	1.26	−1.18	0.0135
Average power (W/kg)	0.210	1.23	3.01	0.0164
Power decrement (%)	−0.204	0.82	0.20	0.0028
PT/W Flex 180°/s L (Nm/kg)	0.202	1.22	2.21	0.0075

Table 5. *Cont.*

Feature	LR β	Odds Ratio	LDA Coef.	Mean |SHAP|
PT/W Flex 60°/s L (Nm/kg)	0.171	1.19	−0.50	0.0106
PT/W Ext 60°/s L (Nm/kg)	0.168	1.18	0.92	0.0080

Note. β = standardised logistic regression coefficient; OR = odds ratio (e^{β}); LDA coef. = linear discriminant analysis coefficient; SHAP = SHapley Additive exPlanations, computed on the random forest model; values are on the probability scale for the professional class and were computed for the training-set observations; mean |SHAP| = global mean absolute SHAP value. The full 24-feature ranking is available from the corresponding author on request. LR coefficients are penalised (L2) estimates; odds ratios therefore express shrunken per-SD associations and should not be read as unpenalised effect estimates. LDA coefficients are unpenalised and sensitive to residual collinearity (see text).

3.5. Domain-Restricted Model Comparisons

Table 6 and Figure 8 show the domain-restricted model comparisons, i.e., the standalone discrimination of each restricted feature set. Every domain subset discriminated well above chance: anthropometric/training-history alone reached CV-AUC 0.95–0.97; jump-plus-Wingate-power-only, 0.95–0.96; sprint-plus-agility-only, 0.97–0.98; and CMJ+SJ alone, 0.92. Isokinetic PT/W alone was, unexpectedly, the least discriminating standalone subset (CV-AUC 0.77–0.80), and PP[W/kg] alone (often treated as a strong single strength/power predictor in this literature) reached only 0.65–0.73. All subsets combined (24 features) reached CV-AUC 0.998–0.999, exceeding any individual domain. This pattern indicates that discriminatory information is distributed across several domains rather than concentrated in a single physical quality. Standalone subset AUCs do not, however, demonstrate interactions or physiological synergy between domains, nor the incremental contribution of each domain to the full model. Because the subsets differ in size (1–8 features), subsets of equal size were also compared directly: among the eight-feature subsets, isokinetic PT/W (CV-AUC 0.765–0.802) remained clearly weaker than jump plus Wingate power (0.951–0.959), and among the four-feature subsets, sprint and agility (0.972–0.980) matched or exceeded anthropometry/training history (0.955–0.966). Two performance-only subsets (jump plus Wingate power; sprint and agility) therefore discriminated almost completely without any anthropometric or training-history variable.

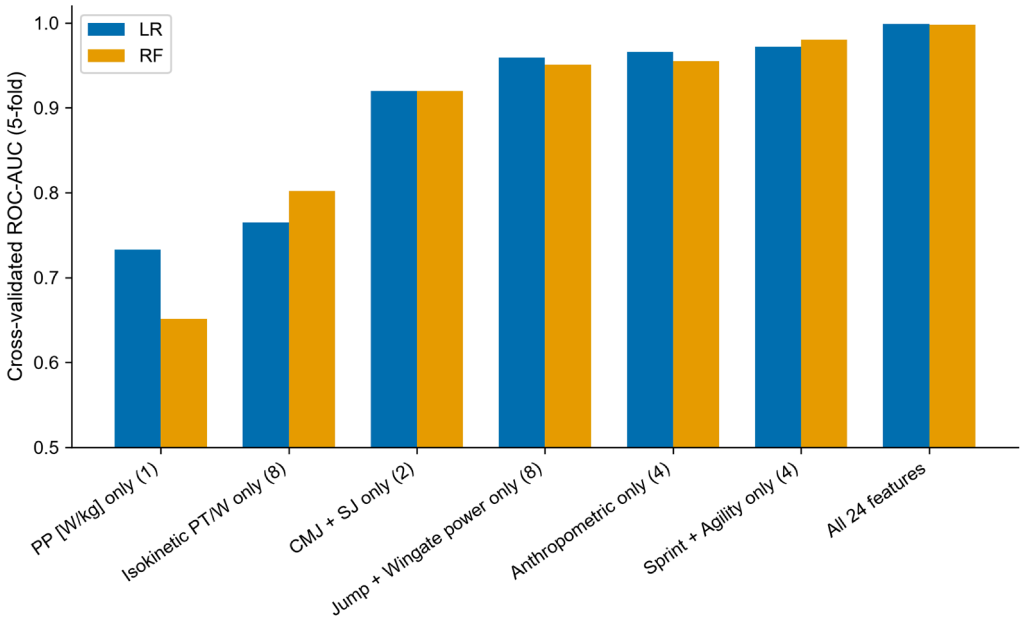


Figure 8. Domain-restricted model comparisons: cross-validated discrimination by feature subset.

Table 6. Cross-validated (5-fold) performance of domain-restricted feature subsets.

Feature Subset	N Features	CV-AUC (LR)	CV-AUC (RF)
PP [W/kg] only (1)	1	0.733	0.651
Isokinetic PT/W only (8)	8	0.765	0.802
CMJ + SJ only (2)	2	0.920	0.920
Jump + Wingate power only (8)	8	0.959	0.951
Anthropometric only (4)	4	0.966	0.955
Sprint + Agility only (4)	4	0.972	0.980
All 24 features	24	0.999	0.998

Note. CV-AUC = 5-fold cross-validated area under the receiver operating characteristic curve; LR = logistic regression; RF = random forest; PP = peak power; CMJ = countermovement jump; SJ = squat jump. Values are point estimates from a single 5-fold CV on the full sample; differences of 0.02 or less between subsets were not interpreted. The jump + Wingate subset combines vertical-jump/leg-stiffness and cycle-ergometer variables and is not a jump-derived measure.

3.6. Multivariate Visualisation

Figure 9 shows the two-dimensional PCA projection of the 24 standardised features (fitted on the full sample for visualisation only; PCA did not inform model training). The first two components explained 37.5% and 14.8% of total variance, respectively, and amateur and professional players separated visibly along the first principal component (PC1), consistent with the near-ceiling classification performance obtained from the full feature set.

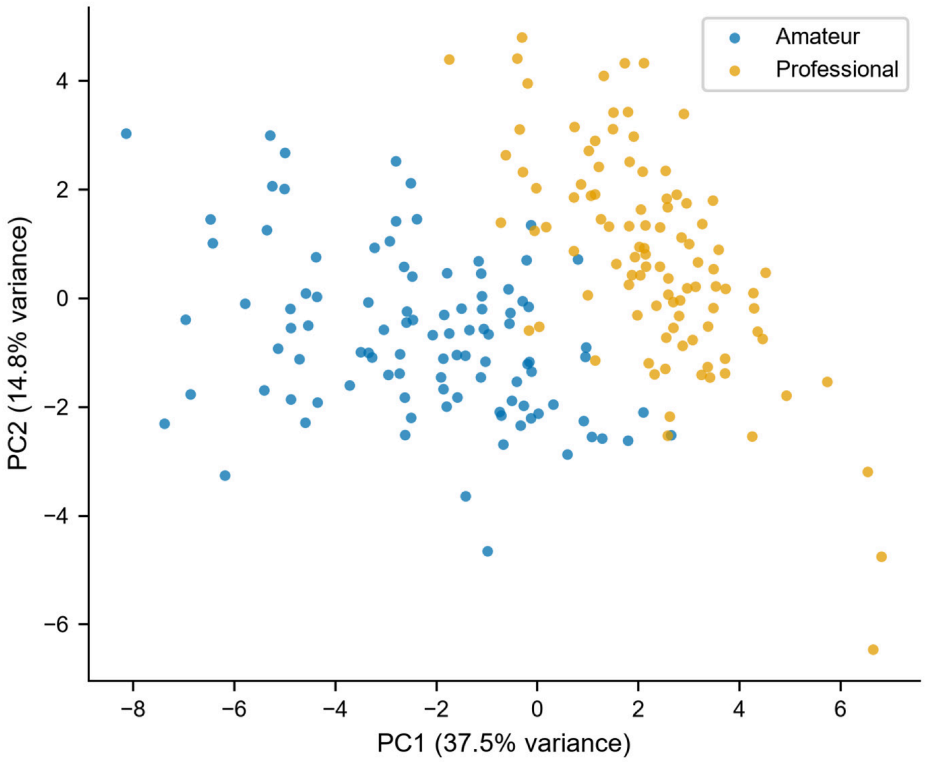


Figure 9. PCA projection of the 24-feature performance profile.

4. Discussion

This study examined amateur-versus-professional discrimination in soccer players using the full available physical performance battery rather than a single domain, applying five classifiers alongside linear and non-linear explainable machine learning. Two results stand out. First, when anthropometry/training history, isokinetic strength, vertical-jump

and elastic-energy indices, Wingate anaerobic power, reactive agility and sprint performance are combined, discrimination between amateur and professional players is essentially at ceiling (AUC 0.998–1.000). Second, this near-ceiling performance is not carried by any single quality: training history, vertical-jump and Wingate power, sprint and reactive agility each discriminate strongly on their own, and isokinetic relative strength, the domain most often relied on in this literature, was the least discriminating standalone domain in the domain-restricted comparisons (Table 6) when torque was expressed relative to body mass; this concerns discriminatory value in these two groups, not the physiological importance of strength.

The tested hypotheses were largely supported, within the limits of a cross-sectional comparison. Domains beyond isokinetic strength contributed substantially to discrimination (hypothesis i): relative strength and power remained significant and meaningful (consistent with [2–5,7,8,10]), but sprint, agility and training history contributed more in this cohort, aligning with literature that emphasises acceleration and change-of-direction capacity as central to competitive-level separation in soccer [25,26,28,30–33] and with reports that training history and years of structured practice track competitive attainment [83,84]. Relative isokinetic strength and power remained significant but were not the dominant domain once the fuller battery was modelled (hypothesis ii; Table 6). Training age reflects accumulated exposure to organised training and the selection processes through which players attain professional status, so classification driven by training history must be distinguished from discrimination based on physical capacities. The more conservative test of hypothesis (i) is therefore provided by the performance-only analyses: the jump-plus-Wingate-power and sprint-plus-agility subsets reached CV-AUC 0.95–0.98 without any anthropometric or training-history variable (Table 6). SHAP importance corroborated the univariate and domain-restricted evidence and, once the deterministic-redundancy variables (H:Q ratio, CMJ–SJ difference, SSC%) were excluded from the multivariable model, coefficient-based and SHAP-based rankings converged with the univariate effect-size ranking (hypothesis iii). Whether entering a derived index alongside its components is problematic depends on the type of dependence and on the model: an exact linear combination (the CMJ–SJ difference) makes linear models non-identifiable, whereas nonlinear indices such as SSC% and the H:Q ratio [11,12,14,15] create strong but model-specific dependence; neither case is in itself evidence of suppression [48,54].

The magnitude of the discrimination obtained here needs careful interpretation. Three lines of evidence indicate that the near-ceiling AUC is stable within this dataset and not attributable to a single favourable split or a gross pipeline error: nested cross-validation and an internal stability assessment with 20 repetitions of stratified five-fold CV (100 non-independent folds) showed narrow spread (SD $\leq$ 0.005); a label-permutation test placed the observed AUC far outside a null distribution centred at chance ($p = 0.001$); and an unsupervised, unfitted composite index of the same 24 features already separated the groups almost as well (AUC = 0.969) as the tuned classifiers, indicating that the separation does not depend on any specific algorithm. None of these checks, however, can exclude systematic differences between the recruited groups, such as club-specific training and testing conditions, recruitment sources, season phase or assessor effects, which are shared by players in the training and test partitions under player-level random splitting [85]. Because competitive status is nested within clubs, club-grouped validation would produce held-out folds containing a single class, so this source of optimism can be addressed only by external validation in independent clubs. The near-complete separation is nonetheless plausible: previous work has shown that professional players differ from amateurs in several physical qualities at once [6,83,86,87]. When many correlated-but-not-redundant qualities each shift consistently in the same direction, their combination can separate groups far more sharply

than any one quality alone; this property of multivariate distance metrics is well documented outside sport [67] and is reflected here in a model-implied Mahalanobis distance of D ≈ 4 (Section 3.3). Nonetheless, the result was obtained in a single cross-sectional cohort, and even after feature curation the ratio of features to training-set observations (24:140) is modest for classifiers as flexible as random forest and gradient boosting. Near-ceiling performance in one sample does not show that the same battery would separate amateur and professional players this cleanly in an independent cohort, league, or country, and external validation is needed before the specific AUC values are treated as generalisable [71–73]. Finally, penalised logistic regression and LDA matched the non-linear ensembles, indicating that the two groups are close to linearly separable in this feature space. Consistent with evidence that flexible machine-learning models rarely outperform logistic regression on tabular data [88,89], the simpler linear model is preferable for interpretation and transfer, pending the external validation that all prediction models require [90].

Methodologically, the study illustrates a caution that is well documented in the multicollinearity literature: under collinearity, multivariate coefficients are not a safe guide to feature importance [48–52]. SSC% and the CMJ–SJ difference were statistically indistinguishable between the two groups on their own (Table 1, both n.s.), yet SSC% is a deterministic nonlinear function of CMJ and SJ height, and the CMJ–SJ difference is their exact linear combination. Had either been entered into the linear models alongside CMJ and SJ height, the coefficients would have been non-identifiable (CMJ–SJ difference) or highly unstable (SSC%). For this reason, both were excluded from the curated 24-feature model and retained only for univariate reporting (Table 1). A related, although nonlinear, dependence applies to the H:Q ratio, a quotient of the flexion and extension PT/W values already in the model. The practical implication is model-specific rather than a blanket rule: when a derived index and its components are candidate predictors, analysts should decide which representation reflects the construct of interest, check identifiability and collinearity for the intended model, and avoid interpreting the coefficients of algebraically dependent variables in isolation [53,54,91]. SHAP, applied to the curated model, agreed closely with univariate effect sizes and domain-restricted results, although SHAP attributions are themselves sensitive to feature dependence [56,57] and the unpenalised LDA coefficients showed sign reversals for three features (Section 3.4). Both observations support the recommendation that explainable-AI outputs be triangulated against classical diagnostics rather than trusted in isolation, particularly in small, feature-rich sports-science datasets [55,71].

The null findings for the CMJ–SJ difference and SSC% should also be read in light of the unusually small CMJ–SJ difference in this cohort (≈0.6–0.8 cm, ≈2%; Table 1) and its large between-player variability (SSC% SD 3.0–5.3%), which implies that many players jumped as high in the SJ as in the CMJ or higher. Small, unintended countermovements during the SJ are common and can raise SJ height by up to 6 cm unless controlled with force–time data [18]; although SJ trials with a visible countermovement were repeated (Section 2.2), visual inspection cannot exclude small-amplitude countermovements, which would compress the CMJ–SJ difference and attenuate any true group difference in stretch-shortening-cycle utilisation. Moreover, the conventional EUR is the CMJ:SJ ratio [16], whereas the variable analysed here is the absolute CMJ–SJ difference in centimetres; the null result therefore concerns the absolute pre-stretch augmentation rather than the ratio, and the physiological interpretation of both derived stretch-shortening-cycle indices should remain cautious.

Body-size normalisation also shapes the domain ranking. Expressed in absolute terms, isokinetic peak torque differed between groups about as much as sprint and agility did (d = 1.41–1.86), whereas body-mass-normalised torque differed less (d = 0.70–0.97); much of the absolute advantage therefore reflects the professionals' greater body mass (d = 1.59), which

entered the model as a separate feature. Dividing torque by body mass is the theoretically appropriate normalisation for torque under geometric similarity(performance = $a \cdot m^b$): Jaric et al. [92] propose b ≈ 1 for joint torque and b ≈ 0.67 for force and power. The exponents were not estimated in this cohort, and whether allometric normalisation would increase or decrease group discrimination for any domain is an empirical question that requires the sensitivity analysis listed among the limitations. The finding that isokinetic strength was the least discriminating standalone domain should therefore be read as applying to body-mass-normalised concentric torque at the tested velocities. Weaker discrimination by normalised torque does not imply that strength is physiologically less important for soccer performance, because domain rankings depend on normalisation, measurement choice and the characteristics of the two groups compared.

For practitioners, the findings suggest that physical-profile monitoring intended to describe differences between competitive levels should not rely on isokinetic testing alone, however well-established that method is in this literature [4,5,7–10]. Training history, vertical-jump height and power, sprint acceleration/maximal-velocity splits, and reactive agility each carried comparable or greater discriminative value than relative isokinetic strength in this cohort. Several of these (jump height, sprint time, reactive agility) can be assessed with simple field tests rather than laboratory isokinetic dynamometry, which may make broad physical-profile monitoring more scalable in grassroots and academy settings where dynamometry access is limited [25,28,30,31]. The present study did not examine injury outcomes, so no inference can be drawn about the value of isokinetic or H:Q-ratio testing for injury-risk screening, a distinct purpose addressed by a separate and itself equivocal literature [11–15]. These implications concern group-level description rather than individual benchmarking, monitoring or talent selection: the observed distributions reflect established adult amateur and professional players in one cohort, measurement reliability in this setting was not quantified, and the data cannot indicate whether an individual player is capable of progressing to professional level. Cross-sectional separation of established amateurs and professionals is far easier than prospective prediction: in talented youth cohorts, physical tests have shown small-to-moderate prognostic value [38], anthropometric and physical profiling did not predict professional contracts over ten years in an elite academy [39], and machine-learning models of selection have reached only moderate discrimination (AUC ≈ 0.69–0.76) [93,94]. The present battery should therefore not be used to select or deselect individual players without prospective, external validation. External validation in other cross-sectional samples from different clubs, leagues, age categories and seasons would establish the transportability of the group discrimination but not, by itself, prospective talent-selection value, which requires longitudinal designs.

Several limitations should be noted. The design remains cross-sectional and associative; it cannot establish whether the identified qualities cause professional attainment, result from years of professional-level training and selection, or both. Training age itself, the single largest effect in the battery, plausibly reflects selection history as much as a trainable physical capacity. Only male players were included, as specified in the approved protocol, to avoid confounding by sex-related physiological differences; a sex- or gender-based analysis was therefore not possible, and the findings may not generalise to female athletes [95], youth categories, or other leagues and countries [83]. Competitive level was treated as a binary amateur/professional split rather than the finer-grained tiers (e.g., academy, sub-elite, senior international) shown elsewhere to carry additional information [6,83,87]. Although curation removed the deterministic dependencies, CMJ and SJ height remained strongly collinear (VIF 21.4 and 20.1), the correlation-based part of the curation used the full sample, and the feature-to-training-sample ratio remains modest for the non-linear classifiers used (≈2.9 events per candidate predictor in the development set, well below the minimum

suggested by Riley et al. [76]), and the near-ceiling AUC obtained here should be treated as a cohort-specific upper bound pending external validation rather than a generally expected effect size. This is also a secondary analysis of data collected during routine performance testing: reliability statistics specific to this dataset are not reported (published values are cited in Section 2.2), tester blinding to competitive level is rarely feasible in such testing, and differences in season phase between the groups, if present, could have influenced the results. Players were clustered within clubs, and player-level random splitting may place players who share club-specific training and testing conditions in both the development and evaluation sets; club membership was not retained in the anonymised dataset, so the number of clubs could not be reported and club-grouped validation was not possible. The stability of the SHAP rankings on held-out observations and the effect of repeating the correlation screening within training folds were also not assessed. The battery did not include aerobic or intermittent-endurance and repeated-sprint tests, which also distinguish professionals from amateurs [35,36], or eccentric (functional) H:Q ratios [15]. Playing position was not considered, so an uneven positional composition of the groups could have contributed to the differences [7,8,83]. Finally, a formal performance-only multivariable model, a performance-only composite index, a leave-one-domain-out analysis with repeated cross-validation, and allometric sensitivity analyses were not performed and remain to be tested.

5. Conclusions

Profiling amateur and professional soccer players across the full available battery of physical performance domains, not isokinetic strength testing alone, achieved near-ceiling discrimination in this cohort (AUC 0.998–1.000) that was stable under nested cross-validation and internal resampling checks. Discriminatory information was distributed across training history, vertical-jump and Wingate power, sprint and reactive agility; body-mass-normalised isokinetic torque was the least discriminating standalone domain, which reflects its discriminatory value in these groups rather than the physiological importance of strength. The near-null univariate discrimination of the CMJ–SJ difference and stretch-shortening-cycle contribution, despite their algebraic dependence on countermovement- and squat-jump height, illustrates why derived indices and their components require model-specific handling: exact or near-deterministic dependence can make multivariable coefficients non-identifiable or unstable. Once these derived variables were excluded, SHAP attributions converged with univariate and domain-restricted evidence, although they describe contributions to classification rather than causal importance. In sports-science machine learning, inference is more robust when several diagnostic approaches are triangulated than when any single one is relied on. Because the design is cross-sectional and single-cohort, these findings describe the separation of established groups; external validation across clubs, leagues, age categories and seasons is required before the approach is used for individual assessment, and prospective designs are needed to establish talent-selection value.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/sports14110521/s1, Table S1: TRIPOD+AI checklist.

Author Contributions: Conceptualization, G.D., F.B., A.G.A. and E.T.; methodology, G.D., F.B. and A.G.A.; software, G.D., F.B. and A.G.A.; validation, G.D., F.B., A.G.A., M.Ç., H.K. and E.T.; formal analysis, G.D., F.B. and A.G.A.; investigation, G.D., F.B. and A.G.A.; resources, G.D., F.B., A.G.A., M.Ç., H.K. and E.T.; data curation, G.D., F.B., A.G.A., M.Ç. and H.K.; writing—original draft preparation, G.D., F.B., A.G.A. and E.T.; writing—review and editing, G.D., F.B., A.G.A. and E.T.; visualization, G.D., F.B. and A.G.A.; supervision, G.D., F.B., A.G.A. and E.T.; project administration, G.D., F.B., A.G.A., M.Ç., H.K. and E.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Commission of Gazi University (protocol code 2026-1573, date of approval 1 September 2026; meeting no. 14; approval letter E-77082166-604.01-1626090 dated 7 September 2026).

Informed Consent Statement: Informed consent for participation was obtained from all subjects involved in the study.

Data Availability Statement: The data and analysis code presented in this study are available on request from the corresponding author.

Acknowledgments: During the preparation of this manuscript, the authors used Claude (Claude Opus 5.5; Anthropic, San Francisco, CA, USA) for the purposes of language editing, adapting the manuscript to the journal's formatting requirements, condensing the abstract, and formatting the reference list and checking bibliographic details (including DOIs) against Crossref. The tool was not used for study design, data collection, statistical analysis or interpretation of the results. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

Figure A1 shows the analytical workflow and the observations used in each analysis. Table A1 reconciles the 45 candidate variables with the 24 features entered into the multivariable models and gives the reason for each exclusion.

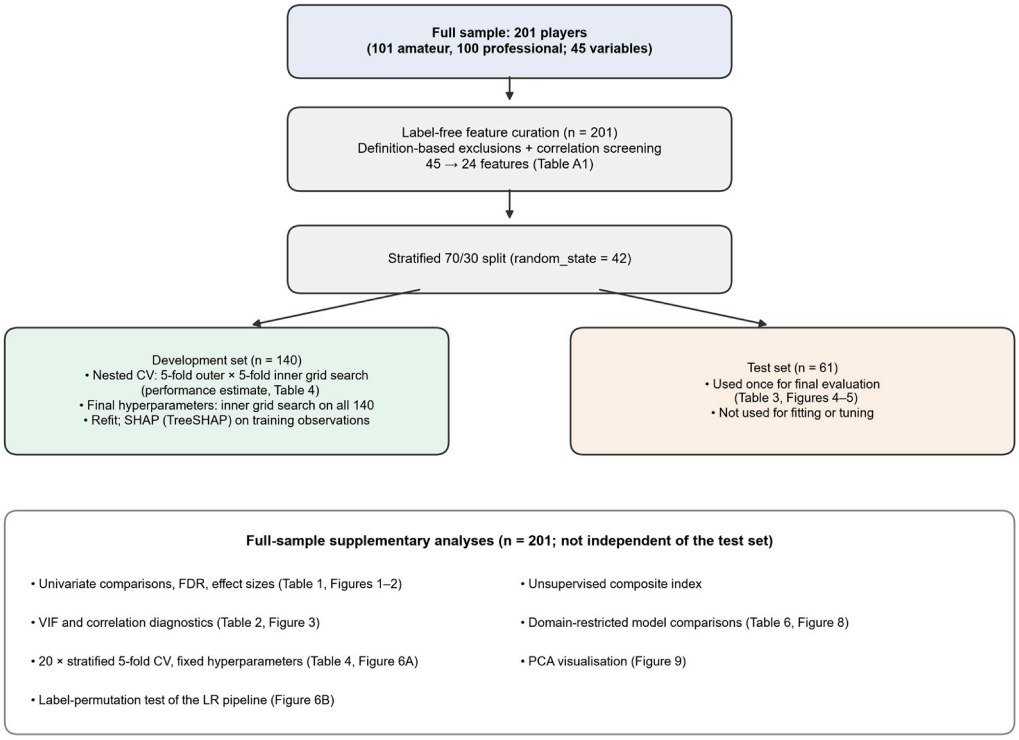


Figure A1. Analytical workflow. The test set ($n = 61$) was used once for final evaluation; feature curation and the supplementary analyses used the full sample ($n = 201$) and are therefore not independent of the test set.

Table A1. Data dictionary of the 45 candidate variables and their status in the multivariable models.

Domain	Variable (s) (Unit) [n]	Definition/Derivation	Status in Multivariable Models
Anthropometry/training history	Age (years) [1]	Chronological age at testing	Retained
	Training age (years) [1]	Self-reported years of organised, club-based soccer training	Retained
	Height (cm) [1]	Stature, wall-mounted stadiometer	Retained
	Body mass (kg) [1]	Calibrated digital scale	Retained
Isokinetic strength, absolute	PT Ext and Flex, 60°/s and 180°/s, L and R (Nm) [8]	Highest gravity-corrected concentric peak torque at each velocity and side	Excluded (definition-based): absolute value; body-mass-normalised equivalent retained and body mass entered separately
Isokinetic strength, relative	PT/W Ext and Flex, 60°/s and 180°/s, L and R (Nm/kg) [8]	Peak torque divided by body mass	Retained (all 8)
H:Q ratio	H:Q ratio, 60°/s and 180°/s, L and R (%) [4]	100 × PT/W Flex/PT/W Ext, same velocity and side	Excluded (definition-based): deterministic nonlinear (quotient) function of retained PT/W values
Vertical jump/leg stiffness	CMJ height (cm) [1]	Flight-time method (OptoJump), best of three trials	Retained
	SJ height (cm) [1]	Flight-time method from a ≈ 90° static semi-squat, best of three trials	Retained
	CMJ–SJ difference (cm) [1]	CMJ − SJ (labelled "EUR" in the original dataset; not the conventional CMJ:SJ ratio)	Excluded (definition-based): exact linear combination of CMJ and SJ
	SSC% [1]	100 × (CMJ − SJ)/SJ	Excluded (definition-based): deterministic nonlinear function of CMJ and SJ
	Leg stiffness (kN/m) [1]	Repeated hopping; spring-mass equation of Dalleau et al. [58]	Retained
Wingate anaerobic power	Peak, average and minimum power (W) [3]	30 s Wingate test; highest, mean and lowest power output	Excluded (definition-based): absolute value; body-mass-normalised equivalent retained and body mass entered separately
	Peak, average and minimum power (W/kg) [3]	The same values divided by body mass	Retained (all 3)
	Time to peak power (ms) [1]	Time from test start to peak power	Retained

Table A1. *Cont.*

Domain	Variable (s) (Unit) [n]	Definition/Derivation	Status in Multivariable Models
	Power decrement (%) [1]	100 × (PP − MP)/PP	Retained: fatigue construct of interest; nonlinear quotient with low linear dependence on retained variables (VIF = 1.7)
Agility	Reactive agility time (s) [1]	Total SpeedCourt test time	Retained
	Mean 180° turn time (s) [1]	Mean of left and right turn times	Retained
	Mean left- and right-turn time (s) [2]	Mean turn time for each direction (SpeedCourt)	Excluded (sample-correlation-based): summarised by mean 180° turn time
Sprint	5 m and 30 m sprint time (s) [2]	Photocell split times, standing start, best of two trials	Retained
	10 m and 20 m sprint time (s) [2]	Photocell split times	Excluded (sample-correlation-based): r = 0.78–0.93 with the retained 5 m and 30 m splits

Note. *n* = number of variables (total 45; 24 retained). Definition-based exclusions do not depend on the sample; sample-correlation-based exclusions used correlations estimated in the full sample (Section 2.4). PT = peak torque; PT/W = peak torque normalised to body mass; Ext = extension; Flex = flexion; L = left; R = right; H:Q = hamstring:quadriceps; CMJ = countermovement jump; SJ = squat jump; SSC% = stretch-shortening-cycle contribution; PP = peak power; MP = minimum power.

References

1. Stølen, T.; Chamari, K.; Castagna, C.; Wisløff, U. Physiology of soccer: An update. *Sports Med.* **2005**, *35*, 501–536. [CrossRef] [PubMed]
2. Wisløff, U.; Castagna, C.; Helgerud, J.; Jones, R.; Hoff, J. Strong correlation of maximal squat strength with sprint performance and vertical jump height in elite soccer players. *Br. J. Sports Med.* **2004**, *38*, 285–288. [CrossRef] [PubMed]
3. Suchomel, T.J.; Nimphius, S.; Stone, M.H. The importance of muscular strength in athletic performance. *Sports Med.* **2016**, *46*, 1419–1449. [CrossRef] [PubMed]
4. Cometti, G.; Maffiuletti, N.A.; Pousson, M.; Chatard, J.C.; Maffulli, N. Isokinetic strength and anaerobic power of elite, subelite and amateur French soccer players. *Int. J. Sports Med.* **2001**, *22*, 45–51. [CrossRef] [PubMed]
5. Bekris, E.; Pidoulas, G.; Pidoulas, P.; Gissis, I.; Katis, A.; Komsis, S. Examination of physical fitness parameters between professional and amateur Greek soccer players during the transition period. *J. Strength Cond. Res.* **2021**, *35*, 776–781. [CrossRef] [PubMed]
6. Beato, M.; Young, D.; Stiff, A.; Coratella, G. Lower-limb muscle strength, anterior-posterior and inter-limb asymmetry in professional, elite academy and amateur soccer players. *J. Hum. Kinet.* **2021**, *77*, 135–146. [CrossRef] [PubMed]
7. Śliwowski, R.; Grygorowicz, M.; Hojszyk, R.; Jadczak, Ł. The isokinetic strength profile of elite soccer players according to playing position. *PLoS ONE* **2017**, *12*, e0182177. [CrossRef] [PubMed]
8. Ruas, C.V.; Minozzo, F.; Pinto, M.D.; Brown, L.E.; Pinto, R.S. Lower-extremity strength ratios of professional soccer players according to field position. *J. Strength Cond. Res.* **2015**, *29*, 1220–1226. [CrossRef] [PubMed]
9. AlTaweel, A.; Nuhmani, S.; Ahsan, M.; Abualait, T.; Muaidi, Q. Determining the hip joint isokinetic muscle strength and range of motion of professional soccer players based on their field position. *PeerJ* **2022**, *10*, e14000. [CrossRef] [PubMed]
10. Kabaciński, J.; Szozda, P.; Mackala, K.; Murawa, M.; Rzepnicka, A.; Szewczyk, P.; Dworak, L. Relationship between isokinetic knee strength and speed, agility, and explosive power in elite soccer players. *Int. J. Environ. Res. Public Health* **2022**, *19*, 671. [CrossRef] [PubMed]
11. Dauty, M.; Menu, P.; Fouasson-Chailloux, A. Cutoffs of isokinetic strength ratio and hamstring strain prediction in professional soccer players. *Scand. J. Med. Sci. Sports* **2018**, *28*, 276–281. [CrossRef] [PubMed]
12. Lee, J.W.Y.; Mok, K.M.; Chan, H.C.K.; Yung, P.S.H.; Chan, K.M. Eccentric hamstring strength deficit and poor hamstring-to-quadriceps ratio are risk factors for hamstring strain injury in football: A prospective study of 146 professional players. *J. Sci. Med. Sport* **2018**, *21*, 789–793. [CrossRef] [PubMed]

13. Freckleton, G.; Pizzari, T. Risk factors for hamstring muscle strain injury in sport: A systematic review and meta-analysis. *Br. J. Sports Med.* **2013**, *47*, 351–358. [CrossRef] [PubMed]
14. Baroni, B.M.; Ruas, C.V.; Ribeiro-Alvares, J.B.; Pinto, R.S. Hamstring-to-quadriceps torque ratios of professional male soccer players: A systematic review. *J. Strength Cond. Res.* **2020**, *34*, 281–293. [CrossRef] [PubMed]
15. Ruas, C.V.; Pinto, R.S.; Haff, G.G.; Lima, C.D.; Pinto, M.D.; Brown, L.E. Alternative methods of determining hamstrings-to-quadriceps ratios: A comprehensive review. *Sports Med. Open* **2019**, *5*, 11. [CrossRef] [PubMed]
16. McGuigan, M.R.; Doyle, T.L.; Newton, M.J.; Edwards, D.; Nimphius, S.; Newton, R. Eccentric utilization ratio: Effect of sport and phase of training. *J. Strength Cond. Res.* **2006**, *20*, 992–995. [CrossRef] [PubMed]
17. Markovic, G.; Mikulic, P. Neuro-musculoskeletal and performance adaptations to lower-extremity plyometric training. *Sports Med.* **2010**, *40*, 859–895. [CrossRef] [PubMed]
18. Van Hooren, B.; Zolotarjova, J. The difference between countermovement and squat jump performances: A review of underlying mechanisms with practical applications. *J. Strength Cond. Res.* **2017**, *31*, 2011–2020. [CrossRef] [PubMed]
19. Radnor, J.M.; Oliver, J.L.; Waugh, C.M.; Myer, G.D.; Moore, I.S.; Lloyd, R.S. The influence of growth and maturation on stretch-shortening cycle function in youth. *Sports Med.* **2018**, *48*, 57–71. [CrossRef] [PubMed]
20. Suchomel, T.J.; Sole, C.J.; Stone, M.H. Comparison of methods that assess lower-body stretch-shortening cycle utilization. *J. Strength Cond. Res.* **2016**, *30*, 547–554. [CrossRef] [PubMed]
21. Kilci, A.; Korkmaz Eryılmaz, S.G.; Boyraz, Ö.C.; Koç, M.E.; Binokay, H.; Kamiş, O.; Praça, G.M. Exploring the relationship between reactive strength index, leg stiffness, and physical performance in adolescent male soccer players: The mediating role of maturation. *Int. J. Sports Sci. Coach.* **2026**, *21*, 832–842. [CrossRef]
22. Jarvis, P.; Turner, A.; Read, P.; Bishop, C. Reactive strength index and its associations with measures of physical and sports performance: A systematic review with meta-analysis. *Sports Med.* **2022**, *52*, 301–330. [CrossRef] [PubMed]
23. Cerrah, A.; Onarici Gungor, E.; Soylu, A.; Ertan, H. Muscular activation differences between professional and amateur soccer players during countermovement jump. *Turk. J. Sport Exerc.* **2014**, *16*, 51–58. [CrossRef]
24. Castagna, C.; Castellini, E. Vertical jump performance in Italian male and female national team soccer players. *J. Strength Cond. Res.* **2013**, *27*, 1156–1161. [CrossRef] [PubMed]
25. Haugen, T.; Tønnessen, E.; Hisdal, J.; Seiler, S. The role and development of sprinting speed in soccer. *Int. J. Sports Physiol. Perform.* **2014**, *9*, 432–441. [CrossRef] [PubMed]
26. Morin, J.-B.; Gimenez, P.; Edouard, P.; Arnal, P.; Jiménez-Reyes, P.; Samozino, P.; Brughelli, M.; Mendiguchia, J. Sprint acceleration mechanics: The major role of hamstrings in horizontal force production. *Front. Physiol.* **2015**, *6*, 404. [CrossRef] [PubMed]
27. López-Segovia, M.; Marques, M.C.; van den Tillaar, R.; González-Badillo, J.J. Relationships between vertical jump and full squat power outputs with sprint times in U21 soccer players. *J. Hum. Kinet.* **2011**, *30*, 135–144. [CrossRef] [PubMed]
28. Bergkamp, T.L.G.; Niessen, A.S.M.; den Hartigh, R.J.R.; Frencken, W.G.P.; Meijer, R.R. Methodological issues in soccer talent identification research. *Sports Med.* **2019**, *49*, 1317–1335. [CrossRef] [PubMed]
29. Kutlu, M.; Yapıcı, H.; Yoncalık, O.; Çelik, S. Comparison of a new test for agility and skill in soccer with other agility tests. *J. Hum. Kinet.* **2012**, *33*, 143–150. [CrossRef] [PubMed]
30. Pojskic, H.; Åslin, E.; Krolo, A.; Jukic, I.; Uljevic, O.; Spasic, M.; Sekulic, D. Importance of reactive agility and change of direction speed in differentiating performance levels in junior soccer players: Reliability and validity of newly developed soccer-specific tests. *Front. Physiol.* **2018**, *9*, 506. [CrossRef] [PubMed]
31. Altmann, S.; Ringhof, S.; Neumann, R.; Woll, A.; Rumpf, M.C. Validity and reliability of speed tests used in soccer: A systematic review. *PLoS ONE* **2019**, *14*, e0220982. [CrossRef] [PubMed]
32. Little, T.; Williams, A.G. Specificity of acceleration, maximum speed, and agility in professional soccer players. *J. Strength Cond. Res.* **2005**, *19*, 76–78. [CrossRef] [PubMed]
33. Kaplan, T.; Erkmen, N.; Taskin, H. The evaluation of the running speed and agility performance in professional and amateur soccer players. *J. Strength Cond. Res.* **2009**, *23*, 774–778. [CrossRef] [PubMed]
34. Haugen, T.A.; Tønnessen, E.; Seiler, S. Anaerobic performance testing of professional soccer players 1995–2010. *Int. J. Sports Physiol. Perform.* **2013**, *8*, 148–156. [CrossRef] [PubMed]
35. Rampinini, E.; Sassi, A.; Morelli, A.; Mazzoni, S.; Fanchini, M.; Coutts, A.J. Repeated-sprint ability in professional and amateur soccer players. *Appl. Physiol. Nutr. Metab.* **2009**, *34*, 1048–1054. [CrossRef] [PubMed]
36. Wells, C.M.; Edwards, A.M.; Winter, E.M.; Fysh, M.L.; Drust, B. Sport-specific fitness testing differentiates professional from amateur soccer players where VO_2max and VO_2 kinetics do not. *J. Sports Med. Phys. Fit.* **2012**, *52*, 245–254.
37. Sarmento, H.; Anguera, M.T.; Pereira, A.; Araújo, D. Talent identification and development in male football: A systematic review. *Sports Med.* **2018**, *48*, 907–931. [CrossRef] [PubMed]
38. Murr, D.; Raabe, J.; Höner, O. The prognostic value of physiological and physical characteristics in youth soccer: A systematic review. *Eur. J. Sport Sci.* **2018**, *18*, 62–74. [CrossRef] [PubMed]

39. Craig, T.P.; Swinton, P. Anthropometric and physical performance profiling does not predict professional contracts awarded in an elite Scottish soccer academy over a ten-year period. *Eur. J. Sport Sci.* **2021**, *21*, 1101–1110. [CrossRef] [PubMed]
40. Rommers, N.; Rössler, R.; Verhagen, E.; Vandecasteele, F.; Verstockt, S.; Vaeyens, R.; Lenoir, M.; D'Hondt, E.; Witvrouw, E. A machine learning approach to assess injury risk in elite youth football players. *Med. Sci. Sports Exerc.* **2020**, *52*, 1745–1751. [CrossRef] [PubMed]
41. Rico-González, M.; Pino-Ortega, J.; Méndez, A.; Clemente, F.M.; Baca, A. Machine learning application in soccer: A systematic review. *Biol. Sport* **2023**, *40*, 249–263. [CrossRef] [PubMed]
42. Rossi, A.; Pappalardo, L.; Cintia, P. A narrative review for a machine learning application in sports: An example based on injury forecasting in soccer. *Sports* **2022**, *10*, 5. [CrossRef] [PubMed]
43. Van Eetvelde, H.; Mendonça, L.D.; Ley, C.; Seil, R.; Tischer, T. Machine learning methods in sport injury prediction and prevention: A systematic review. *J. Exp. Orthop.* **2021**, *8*, 27. [CrossRef] [PubMed]
44. Reis, F.J.J.; Alaiti, R.K.; Vallio, C.S.; Hespanhol, L. Artificial intelligence and machine learning approaches in sports: Concepts, applications, challenges, and future perspectives. *Braz. J. Phys. Ther.* **2024**, *28*, 101083. [CrossRef] [PubMed]
45. Rossi, A.; Pappalardo, L.; Cintia, P.; Iaia, F.M.; Fernàndez, J.; Medina, D. Effective injury forecasting in soccer with GPS training data and machine learning. *PLoS ONE* **2018**, *13*, e0201264. [CrossRef] [PubMed]
46. Barron, D.; Ball, G.; Robins, M.; Sunderland, C. Artificial neural networks and player recruitment in professional soccer. *PLoS ONE* **2018**, *13*, e0205818. [CrossRef] [PubMed]
47. Kelly, A.L.; Williams, C.A.; Cook, R.; Sáiz, S.L.J.; Wilson, M.R. A multidisciplinary investigation into the talent development processes at an English football academy: A machine learning approach. *Sports* **2022**, *10*, 159. [CrossRef] [PubMed]
48. Friedman, L.; Wall, M. Graphical views of suppression and multicollinearity in multiple linear regression. *Am. Stat.* **2005**, *59*, 127–136. [CrossRef]
49. Weaving, D.; Jones, B.; Ireton, M.; Whitehead, S.; Till, K.; Beggs, C.B. Overcoming the problem of multicollinearity in sports performance data: A novel application of partial least squares correlation analysis. *PLoS ONE* **2019**, *14*, e0211776. [CrossRef] [PubMed]
50. Akinwande, M.O.; Dikko, H.G.; Samson, A. Variance inflation factor: As a condition for the inclusion of suppressor variable(s) in regression analysis. *Open J. Stat.* **2015**, *5*, 754–767. [CrossRef]
51. Johnston, R.; Jones, K.; Manley, D. Confounding and collinearity in regression analysis: A cautionary tale and an alternative procedure, illustrated by studies of British voting behaviour. *Qual. Quant.* **2018**, *52*, 1957–1976. [CrossRef] [PubMed]
52. Kraha, A.; Turner, H.; Nimon, K.; Zientek, L.R.; Henson, R.K. Tools to support interpreting multiple regression in the face of multicollinearity. *Front. Psychol.* **2012**, *3*, 44. [CrossRef] [PubMed]
53. Salmerón, R.; García, C.B.; García, J. Variance inflation factor and condition number in multiple linear regression. *J. Stat. Comput. Simul.* **2018**, *88*, 2365–2384. [CrossRef]
54. Ludlow, L.; Klein, K. Suppressor variables: The difference between 'is' versus 'acting as'. *J. Stat. Educ.* **2014**, *22*. [CrossRef]
55. Lundberg, S.M.; Lee, S.-I. A unified approach to interpreting model predictions. In *Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017*; Curran Associates: Red Hook, NY, USA, 2017; pp. 4765–4774.
56. Aas, K.; Jullum, M.; Løland, A. Explaining individual predictions when features are dependent: More accurate approximations to Shapley values. *Artif. Intell.* **2021**, *298*, 103502. [CrossRef]
57. Kumar, I.E.; Venkatasubramanian, S.; Scheidegger, C.; Friedler, S. Problems with Shapley-value-based explanations as feature importance measures. In *Proceedings of the 37th International Conference on Machine Learning (ICML 2020), Virtual Event, 13–18 July 2020*; PMLR: Cambridge, MA, USA, 2020; Volume 119, pp. 5491–5500.
58. Dalleau, G.; Belli, A.; Viale, F.; Lacour, J.R.; Bourdin, M. A simple method for field measurements of leg stiffness in hopping. *Int. J. Sports Med.* **2004**, *25*, 170–176. [CrossRef] [PubMed]
59. Nikolaidis, P.T.; Matos, B.; Clemente, F.M.; Bezerra, P.; Camões, M.; Rosemann, T.; Knechtle, B. Normative data of the Wingate anaerobic test in 1 year age groups of male soccer players. *Front. Physiol.* **2018**, *9*, 1619. [CrossRef] [PubMed]
60. Sheppard, J.M.; Young, W.B. Agility literature review: Classifications, training and testing. *J. Sports Sci.* **2006**, *24*, 919–932. [CrossRef] [PubMed]
61. Düking, P.; Born, D.P.; Sperlich, B. The SpeedCourt: Reliability, usefulness, and validity of a new method to determine change-of-direction speed. *Int. J. Sports Physiol. Perform.* **2016**, *11*, 130–134. [CrossRef] [PubMed]
62. Dirnberger, J.; Wiesinger, H.P.; Kösters, A.; Müller, E. Reproducibility for isometric and isokinetic maximum knee flexion and extension measurements using the IsoMed 2000-dynamometer. *Isokinet. Exerc. Sci.* **2012**, *20*, 149–153. [CrossRef]
63. Markovic, G.; Dizdar, D.; Jukic, I.; Cardinale, M. Reliability and factorial validity of squat and countermovement jump tests. *J. Strength Cond. Res.* **2004**, *18*, 551–555. [CrossRef] [PubMed]
64. Ozkaya, O.; Balci, G.A.; As, H.; Vardarli, E. The test-retest reliability of new generation power indices of Wingate all-out test. *Sports* **2018**, *6*, 31. [CrossRef] [PubMed]

65. Lakens, D. Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Front. Psychol.* **2013**, *4*, 863. [CrossRef] [PubMed]
66. Benjamini, Y.; Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Methodol.* **1995**, *57*, 289–300. [CrossRef]
67. Del Giudice, M.; Booth, T.; Irwing, P. The distance between Mars and Venus: Measuring global sex differences in personality. *PLoS ONE* **2012**, *7*, e29265. [CrossRef] [PubMed]
68. McGraw, K.O.; Wong, S.P. A common language effect size statistic. *Psychol. Bull.* **1992**, *111*, 361–365. [CrossRef]
69. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
70. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [CrossRef]
71. Vabalas, A.; Gowen, E.; Poliakoff, E.; Casson, A.J. Machine learning algorithm validation with a limited sample size. *PLoS ONE* **2019**, *14*, e0224365. [CrossRef] [PubMed]
72. Arlot, S.; Celisse, A. A survey of cross-validation procedures for model selection. *Stat. Surv.* **2010**, *4*, 40–79. [CrossRef]
73. Krstajić, D.; Buturović, L.J.; Leahy, D.E.; Thomas, S. Cross-validation pitfalls when selecting and assessing regression and classification models. *J. Cheminform.* **2014**, *6*, 10. [CrossRef] [PubMed]
74. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
75. Heinze, G.; Schemper, M. A solution to the problem of separation in logistic regression. *Stat. Med.* **2002**, *21*, 2409–2419. [CrossRef] [PubMed]
76. Riley, R.D.; Snell, K.I.E.; Ensor, J.; Burke, D.L.; Harrell, F.E., Jr.; Moons, K.G.M.; Collins, G.S. Minimum sample size for developing a multivariable prediction model: Part II—Binary and time-to-event outcomes. *Stat. Med.* **2019**, *38*, 1276–1296. [CrossRef] [PubMed]
77. DeLong, E.R.; DeLong, D.M.; Clarke-Pearson, D.L. Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics* **1988**, *44*, 837–845. [CrossRef] [PubMed]
78. Bates, S.; Hastie, T.; Tibshirani, R. Cross-validation: What does it estimate and how well does it do it? *J. Am. Stat. Assoc.* **2024**, *119*, 1434–1445. [CrossRef] [PubMed]
79. Brown, L.D.; Cai, T.T.; DasGupta, A. Interval estimation for a binomial proportion. *Stat. Sci.* **2001**, *16*, 101–133. [CrossRef]
80. Collins, G.S.; Moons, K.G.M.; Dhiman, P.; Riley, R.D.; Beam, A.L.; Van Calster, B.; Ghassemi, M.; Liu, X.; Reitsma, J.B.; van Smeden, M.; et al. TRIPOD+AI statement: Updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ* **2024**, *385*, e078378. [CrossRef] [PubMed]
81. Ojala, M.; Garriga, G.C. Permutation tests for studying classifier performance. *J. Mach. Learn. Res.* **2010**, *11*, 1833–1863.
82. Phipson, B.; Smyth, G.K. Permutation p-values should never be zero: Calculating exact p-values when permutations are randomly drawn. *Stat. Appl. Genet. Mol. Biol.* **2010**, *9*, 39. [CrossRef] [PubMed]
83. Slimani, M.; Nikolaidis, P. Anthropometric and physiological characteristics of male soccer players according to their competitive level, playing position and age group: A systematic review. *J. Sports Med. Phys. Fit.* **2019**, *59*, 141–163. [CrossRef] [PubMed]
84. Helsen, W.F.; Starkes, J.L.; Hodges, N.J. Team sports and the theory of deliberate practice. *J. Sport Exerc. Psychol.* **1998**, *20*, 12–34. [CrossRef]
85. Roberts, D.R.; Bahn, V.; Ciuti, S.; Boyce, M.S.; Elith, J.; Guillera-Arroita, G.; Hauenstein, S.; Lahoz-Monfort, J.J.; Schröder, B.; Thuiller, W.; et al. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography* **2017**, *40*, 913–929. [CrossRef]
86. Dellal, A.; Hill-Haas, S.; Lago-Peñas, C.; Chamari, K. Small-sided games in soccer: Amateur vs. professional players' physiological responses, physical, and technical activities. *J. Strength Cond. Res.* **2011**, *25*, 2371–2381. [CrossRef] [PubMed]
87. Rebelo, A.; Brito, J.; Maia, J.; Coelho-e-Silva, M.J.; Figueiredo, A.J.; Bangsbo, J.; Malina, R.M.; Seabra, A. Anthropometric characteristics, physical fitness and technical performance of under-19 soccer players by competitive level and field position. *Int. J. Sports Med.* **2013**, *34*, 312–317. [CrossRef] [PubMed]
88. Christodoulou, E.; Ma, J.; Collins, G.S.; Steyerberg, E.W.; Verbakel, J.Y.; Van Calster, B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *J. Clin. Epidemiol.* **2019**, *110*, 12–22. [CrossRef] [PubMed]
89. Becerra-Patiño, B.A.; Villaseca-Vicuña, R.; Rada-Perdigón, D.A.; Paucar-Uribe, J.D.; Valencia-Sánchez, W.G.; López-Gil, J.F.; Yáñez-Sepúlveda, R. Performance of professional soccer players: Multilevel classification using machine learning and SHAP explainability by playing position. *Data* **2026**, *11*, 190. [CrossRef]
90. Steyerberg, E.W.; Harrell, F.E., Jr. Prediction models need appropriate internal, internal-external, and external validation. *J. Clin. Epidemiol.* **2016**, *69*, 245–247. [CrossRef] [PubMed]
91. Kronmal, R.A. Spurious correlation and the fallacy of the ratio standard revisited. *J. R. Stat. Soc. Ser. A Stat. Soc.* **1993**, *156*, 379–392. [CrossRef]

92. Jaric, S.; Mirkov, D.; Markovic, G. Normalizing physical performance tests for body size: A proposal for standardization. *J. Strength Cond. Res.* **2005**, *19*, 467–474. [CrossRef] [PubMed]
93. Altmann, S.; Ruf, L.; Thiem, S.; Beckmann, T.; Wohak, O.; Romeike, C.; Härtel, S. Prediction of talent selection in elite male youth soccer across 7 seasons: A machine-learning approach. *J. Sports Sci.* **2024**, *42*, 2481–2494. [CrossRef] [PubMed]
94. Jauhiainen, S.; Äyrämö, S.; Forsman, H.; Kauppi, J.P. Talent identification in soccer using a one-class support vector machine. *Int. J. Comput. Sci. Sport* **2019**, *18*, 125–136. [CrossRef]
95. Emmonds, S.; Nicholson, G.; Beggs, C.B.; Jones, B.; Bissas, A. Importance of physical qualities for speed and change of direction ability in elite female soccer players. *J. Strength Cond. Res.* **2019**, *33*, 1669–1677. [CrossRef] [PubMed]

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