©Journal of Sports Science and Medicine (2026) **25**, 58-83
http://www.jssm.org DOI: https://doi.org/10.52082/jssm.2026.58

Review article

The Role of Machine Learning in Talent Identification for Team Sports: A Systematic Review

Qingrong Tang [1], Xiufang Wei [2]✉ and Bo Tan [3]
[1] Geely University of China, Chengdu, China; [2] College of physical education and health, China West Normal University, Nanchong, China; [3] College of physical education and health, Geely University of China, Chengdu, China

Abstract
Talent identification (TID) in team sports is complex, influenced by biological, technical, psychological, and socio-cultural factors. Machine learning (ML) offers tools to integrate high-dimensional data, yet its applications in youth TID remain underexplored. Objectives: To systematically review ML approaches applied to youth talent identification in team sports, with emphasis on data domains, algorithms, validation strategies, and interpretability. Eligible studies included peer-reviewed quantitative research applying ML to youth athletes (≤21 years) in team sports for TID outcomes. Searches were conducted in PubMed, Scopus, and Web of Science, supplemented by reference and citation screening. Extracted data items included input data domains (anthropometric, physical, technical, perceptual–cognitive, psychological, socio-cultural, and multi-domain), ML approach, validation methods, performance metrics (e.g., accuracy, AUC, F1-score), and interpretability techniques. Risk-of-bias assessment was implemented using PROBAST. From 228 records, 27 studies met inclusion criteria. Soccer was most studied (n = 13), with others covering rugby, basketball, cricket, volleyball, and Australian football. Sample sizes ranged from 21 to 13,876 athletes, predominantly male. Supervised algorithms (Random Forest, gradient boosting, neural networks, penalized regression) were most common; some studies used unsupervised clustering. Validation practices varied, with few employing nested cross-validation or external testing. Reported discrimination metrics ranged from modest to excellent (ROC-AUC ≈ 0.58 - 0.96, depending on model and context), yet calibration performance (e.g., Brier score, calibration slope) was rarely reported, and external validation was uncommon. Across studies, predictive accuracy was moderate to high internally but rarely externally confirmed. Risk of bias was high in 59 % of studies, mainly due to inadequate analysis and limited generalizability. Overall, ML shows potential to complement, not replace, traditional TID approaches - acting as a decision-support and hypothesis-generation tool that can assist practitioners in early screening, individualized progression modeling, and evidence-based talent forecasting. To strengthen translational impact, future research should emphasize transparent reporting, calibration assessment, and external validation to ensure robust, applicable ML models for sport talent systems.

Key words: Youth athletes, talent development, predictive modeling, sports analytics, artificial intelligence.

Introduction

Talent identification and development (TID) can be conceptualized as a complex, non-linear, and adaptive system arising from the continuous interaction of multiple constraints (e.g., biological, technical, psychological, environmental, and sociocultural), consistent with ecological dynamics (Vaeyens et al., 2008; Seifert et al., 2017; 2022). This perspective treats athletes and teams as complex adaptive systems in which performance emerges from performer–environment couplings rather than from any single determinant, helping explain variability and divergent pathways to expertise (Seifert et al., 2017; 2022). These same principles inform the use of machine learning (ML), as algorithms trained on representative, context-rich data can better capture the functional - rather than merely descriptive - aspects of performance. Incorporating contextualized variables such as opponent positioning, temporal constraints, or perceptual–motor demands enables ML models to infer how athletes adapt to dynamic environments, thereby aligning data-driven modeling with the ecological validity of real performance contexts (Reis et al., 2024; Cordeiro et al., 2025).

TID outcomes can be operationally as measurable indicators of athlete progression, including selection (the identification or nomination of athletes for higher-level squads, academies, or representative team) (Larkin and O'Connor, 2017), advancement (continued inclusion or promotion within developmental pathways across time); or retention (sustained participation or non-deselection within structured development systems) (Güllich, 2014). TID are central pillars of performance pathways in team sports, yet they remain challenging due to the multifactorial and long-term nature of sporting excellence (Vaeyens et al., 2008). In soccer and other team sports games, early reviews already emphasized that no single anthropometric, physiological, or psychological attribute uniquely determines future elite status, underscoring the need for multidimensional assessment (Williams and Reilly, 2000). Accordingly, comprehensive, multidisciplinary test batteries have been advocated to distinguish performance levels in youth players, integrating technical, physical, and perceptual–cognitive factors (Reilly et al., 2000).

However, conventional selection practices can be biased by structural and developmental factors (Till and Baker, 2020). Across sports, annual age-grouping systematically produces relative age effects that distort participation and attainment, with robust meta-analytic evidence showing substantial over-representation of relatively older athletes (Cobley et al., 2009). These biases also affect women's sport, where relative age effects are prevalent and can shape pathway opportunities (Smith et al., 2018). In parallel, differences in growth and biological maturation are particularly salient in adolescence, where earlier-developing youth may temporarily appear superior in test batteries, complicating prognostic judgments in talent pathways (Malina et al., 2015). Longitudinal work further suggests that while some anthropometric and running measures

Received: 22 October 2025 / Accepted: 20 November 2025 / Published (online): 01 March 2026

show short-term stability, predictability erodes as the follow-up window lengthens, cautioning against early deterministic selection (Deprez et al., 2015).

From a systems perspective, team sports exhibit properties of complex adaptive systems in which performance emerges from interacting constraints across performers, tasks, and environments, challenging linear prediction (Seifert et al., 2017). This lens encourages practitioners to design representative learning environments and assess adaptable skill, rather than isolated traits alone (Woods et al., 2020). Concurrently, the proliferation of player monitoring - such as global positioning system (GPS) and inertial technologies - has generated high-volume, multi-source data that can complement traditional scouting in talent pathways (Ravé et al., 2020). For youth programs in particular, such data-rich approaches may help disentangle transient growth effects from underlying skill and potential, if analyzed with appropriate modeling strategies.

Machine learning (ML) methods are well suited to model high-dimensional, nonlinear relationships and to fuse heterogeneous data streams, and have transformed predictive analytics across biomedicine in analogous problems (Topol, 2019). Within sport, researchers have highlighted the growing role of artificial intelligence (AI) and ML for decision support across performance and recruitment domains (Chmait and Westerbeek, 2021). Indeed, soccer-specific syntheses now document rapid expansion of ML applications, signaling both opportunity and methodological variability that warrant careful appraisal (Rico-González et al., 2023; Beato et al., 2025).

Within this growing landscape, ML applications in TID can be conceptually grouped into four possible overlapping roles. First, predictive modeling seeks to forecast future selection, progression, or performance based on multidimensional athlete data, aligning with conventional supervised learning paradigms (Altmann et al., 2024). Second, clustering and representation learning use unsupervised methods to identify latent groupings or archetypes of players, informing talent grouping and developmental profiling (Contreras-García et al., 2024; Haan et al., 2025). Third, longitudinal monitoring leverages sequential or temporal models to track developmental trajectories and maturation dynamics, offering insight into non-linear growth patterns (Chmait and Westerbeek, 2021). Finally, decision-support systems integrate these analytic layers into practical tools that complement coach judgment by providing interpretable, data-informed recommendations (Chmait and Westerbeek, 2021).

ML applications in youth talent identification are beginning to emerge, directly targeting selection and advancement decisions within academies and development squads (Nassis et al., 2023). Recent work in elite youth soccer used supervised algorithms (e.g., gradient-boosted trees) to predict selection versus de-selection across age groups, identifying contributions from speed, change of direction, countermovement jump, aerobic speed reserve, and technical skill (Altmann et al., 2024). A growing line of inquiry also examines how socio-biological factors, particularly the relative age effect and maturation status, may influence data-driven decision-making (Finnegan et al., 2024). ML offers a means to quantify, and potentially mitigate, these entrenched selection biases - depending on how data are sampled, labelled, and validated - thus serving as a test case for fairness and transparency in predictive modelling (Reis et al., 2024). Multidisciplinary approaches have also combined psychosocial and physiological measures with ML to predict youth rugby union selections, illustrating the value of integrating non-physical determinants (Owen et al., 2022). Beyond supervised prediction, unsupervised learning has been explored to derive role-agnostic player groupings from match running data, offering alternative structures for evaluation and development planning (Haan et al., 2025). At the position-specific level, ML classifiers have been applied to discriminate performance tiers in professional goalkeepers, demonstrating how algorithmic profiling can inform specialized talent evaluation (Jamil et al., 2021).

Yet, translating these advances into dependable youth talent decisions requires vigilance about methodological pitfalls common to prediction research (de Jong et al., 2021). Small sample sizes and inadequate validation inflate estimated performance, highlighting the importance of robust procedures such as nested cross-validation and strict separation of training and testing (Vabalas et al., 2019). Data leakage - through feature selection on the full dataset, reusing individuals across folds, or inadvertent temporal contamination - can markedly overstate model accuracy and undermine reproducibility (Kapoor and Narayanan, 2023). Evaluation must also account for class imbalance and choose metrics judiciously, given differing sensitivities of ROC and precision–recall analyses under skewed outcomes typical of selection tasks (Richardson et al., 2024). For clinical-style prediction problems, independent external validation remains essential to estimate generalizability prior to deployment in new cohorts or clubs (Gallitto et al., 2025). Aligned with broader prediction-model science, contemporary reporting guidance (TRIPOD+AI) and risk-of-bias tools (e.g., PROBAST) provide structured expectations for transparency, reproducibility, and appraisal of ML-based models (Wolff et al., 2019; Collins et al., 2024).

Several narrative and systematic reviews have synthesized traditional and methodological approaches to talent identification (TID) in team sports, but without a specific emphasis on ML techniques and their unique challenges (Barraclough et al., 2022). Other reviews focus on ML in soccer broadly or on injury risk prediction, rather than on youth talent identification across multiple team sports and data modalities (Nassis et al., 2023; Leckey et al., 2025). Likewise, sport-specific talento identification syntheses in football underscore multidimensional determinants but do not evaluate the distinct promises and pitfalls of ML for selection decisions across team sports (Sarmento et al., 2018).

Therefore, the purpose of this systematic review is to map, critically appraise, and synthesize applications of ML to youth talent identification in team sports, with attention to data sources, model classes, validation strategies, interpretability, and risk of bias, consistent with contemporary prediction-model guidance. Conceptually, this review also examines whether multidomain ML models – integrating physical, technical, perceptual–cognitive, and psychosocial indicators - capture developmental potential more effectively than single-domain approaches, thereby addressing how the multidimensional nature of athlete development can be represented within predictive frameworks. Analytically, we also quantify the use of nested versus non-nested cross-validation procedures to provide a transparent overview of model evaluation rigor and guide the reproducibility of the synthesis process. Specifically, we aim to catalog the types of athlete data and ML methods used to predict selection and advancement in team sports, evaluate methodological quality, reporting, and validation practices, summarize model performance, calibration, and generalizability, and identify evidence gaps and practical implications for programs and practitioners seeking to integrate ML into selection and development processes.

Methods

The review was conducted and reported in accordance with PRISMA 2020 recommendations to ensure transparent and reproducible synthesis (Page et al., 2021b). Registration was conducted on OSF (osf.io/yn895; October 15, 2025).

Eligibility criteria

PICO criteria

Studies were considered eligible if they addressed the use of ML methods for TID in team sports. Eligibility was defined using a modified PICO framework as follows:

Population (P): Youth athletes (≤21 years) engaged in organized team sports (e.g., soccer, basketball, rugby, hockey, handball, volleyball, American football, baseball, and other team sports). Studies were eligible regardless of competitive level (grassroots, academy, sub-elite, or elite youth), and no restrictions were imposed on sex. Studies focusing exclusively on adult/professional-only cohorts or on individual sports were excluded. We defined "youth" as athletes ≤21 years to align with established competitive tiers and developmental transition points in team sports. In football and other codes, U21 is the terminal youth category preceding senior squads; research shows that experience and performance at U21 best predict subsequent senior participation compared with earlier youth levels, situating age 21 as the practical boundary of the youth pathway (Herrebrøden and Bjørndal, 2022). More broadly, youth-athlete development reviews describe late adolescence and emerging adulthood (late teens–early 20s) as the period when maturation, psychosocial development, and role transitions converge - precisely the window spanned by the U21 tier - supporting the conceptual placement of ≤21 as the end of the formative, pre-senior phase (Varghese et al., 2022). In studies that included both youth (≤21 years) and adult athletes, inclusion was contingent on whether youth-specific results could be clearly identified or disaggregated.

Intervention/Exposure (I): Application of ML algorithms (supervised, unsupervised, reinforcement, or hybrid approaches) to support talent identification or selection processes (e.g., prediction of selection vs. deselection, progression to higher competitive levels, role-agnostic player clustering, or position-specific profiling in youth athletes). Studies limited to traditional statistical analyses without ML components were excluded.

Comparators (C): Comparator groups were not mandatory. Where applicable, comparators could include traditional scouting, expert coach assessment, or alternative analytic approaches (e.g., regression, rule-based classification).

Outcomes (O): Eligible studies had to report at least one youth TID-related outcome, such as predictive accuracy of selection, identification of key features contributing to progression, classification of athlete profiles, or algorithmic discrimination of performance tiers within youth cohorts. Studies were excluded if ML was applied exclusively to non-TID outcomes (e.g., injury prediction, workload monitoring, or tactical analysis), if ML was applied only in adult/professional samples, or if results were not disaggregated to allow extraction of youth TID-specific findings.

Study design and setting

All quantitative empirical studies employing ML algorithms for TID were included, regardless of design (cross-sectional, longitudinal, retrospective, or prospective). Proof-of-concept studies, validation studies, and applied analyses in real-world settings were all eligible. Qualitative studies, narrative commentaries, editorials, opinion pieces, and reviews were excluded, though their reference lists were screened for potential eligible primary studies.

Report characteristics

Only peer-reviewed journal articles were included to ensure methodological rigor. Grey literature, preprints, conference abstracts, theses, and unpublished reports were excluded due to limitations in methodological detail and peer review. Only studies published in English were considered eligible. No restrictions were placed on the year of publication.

Information sources

The literature search was conducted across three major bibliographic databases to ensure coverage of relevant studies: PubMed, Scopus, and the Web of Science Core Collection. No restrictions were applied with respect to publication year, study design, or participant age at the search stage. The final searches of all databases were completed on October 15, 2025.

To complement the electronic database searches, the reference lists of all studies meeting the eligibility criteria were manually examined to identify additional articles not retrieved in the initial search. Reference lists of previous systematic and narrative reviews relevant to talent identification, sports analytics, or the application of machine learning in sport were also screened. Furthermore, backward and forward citation searches were conducted using the Web of Science Core Collection for all included studies to capture any additional eligible publications.

No study registers, trial registries, organizational repositories, or grey literature sources were searched. Only peer-reviewed journal publications retrieved through the databases and reference list searches were included for screening.

Search strategy

The search strategy was designed to capture all available studies addressing the use of ML for TID in team sports. The strategy combined controlled vocabulary terms and free-text words related to “machine learning,” “artificial intelligence,” and “talent identification” with sport-specific terms, following iterative piloting and refinement to balance sensitivity and specificity. The conceptual structure of the strategy was based on a modified PICO approach, focusing on the population of team sport athletes and the intervention or exposure of machine learning applications for talent identification outcomes. The following search strategy was employed: ("machine learning” OR "artificial intelligence” OR "deep learning" OR "supervised learning" OR "unsupervised learning" OR "neural network*" OR "support vector machine*" OR "random forest*" OR "gradient boosting" OR "learning algorithms" OR "bayesian logistic regression" OR “random forest" OR "random forests" OR "trees" OR "elastic net" OR "ridge" OR "lasso" OR "boosting" OR "predictive modeling") AND (talent* OR "talent identification" OR "talent detection" OR "talent development" OR "player selection" OR "athlete selection" OR "talent promotion") AND ("team sport*" OR "soccer" OR "football" OR "basketball" OR "rugby" OR "handball" OR "volleyball" OR "hockey" OR "baseball" OR "softball" OR "lacrosse" OR "water polo").

Selection process

All records identified through database searching were imported into an Excel sheet, and duplicates were removed prior to screening. Two reviewers independently assessed the eligibility of studies against the predefined inclusion and exclusion criteria in title/abstract screening and then in full-text screening. Disagreements between reviewers were resolved through discussion. The reasons for excluding studies at the full-text stage were documented and reported.

Data collection process

Two reviewers independently extracted data from each study. The extracted information was subsequently compared, and any discrepancies were resolved through discussion. No automation tools or machine learning–based systems were used for data collection. Only information explicitly reported in tables, text, or graphs was included.

Data items

The domain of interest was the performance of machine learning models applied to talent identification in youth team sports. Within this domain, data were extracted on predictive or classification performance metrics reported by each study. These included, where available, overall accuracy, sensitivity (recall), specificity, precision, F1-score, area under the receiver operating characteristic curve (AUC-ROC), and area under the precision–recall curve (AUC-PR). When studies reported multiple metrics, all available values were collected to allow for a comprehensive synthesis.

Other domains included talent-related predictions and classifications such as selection versus deselection, progression to higher competition levels, clustering of players into performance profiles, and position- or role-specific identification. Where studies reported longitudinal prediction outcomes, all time points were collected, and no restrictions were applied to the follow-up period. In cases where results were presented using different analysis strategies (e.g., cross-validation folds, test set performance, external validation), all eligible outcomes were extracted, with priority given to independent test set or external validation results when synthesizing evidence.

No changes were made during the review process to the inclusion or definition of outcome domains. All outcome domains compatible with TID were considered equally relevant at the data extraction stage. However, in the interpretation of findings, external validation performance and transparent reporting of prediction quality were considered most critical, as these outcomes are directly aligned with the review’s objectives of evaluating methodological robustness and generalizability.

In addition to outcomes, other variables were extracted from each study to support subgroup analyses and contextual interpretation. Study characteristics included publication year and country of origin. Participant characteristics comprised sample size, sex distribution, age range, competitive context (e.g., grassroots, academy, or elite youth), and where available, indicators of biological maturation. Sport type was also recorded. Data characteristics included the domain of features used (e.g., anthropometric, physical, technical, perceptual - cognitive, psychosocial, or multi-domain) and the methods of data acquisition (e.g., field-based tests, questionnaires, match-derived tracking data).

Machine learning–related variables included the class of algorithms applied (e.g., supervised, unsupervised, ensemble, deep learning), model development strategies (e.g., feature selection, dimensionality reduction), training and validation procedures (e.g., cross-validation, independent test set, external validation), and performance metrics reported. Where available, reporting of interpretability approaches (e.g., feature importance, SHapley Additive exPlanations, Local Interpretable Model-agnostic Explanations) was also extracted. When information was missing or unclear, we recorded it as “not reported” without making assumptions.

Study risk of bias assessment

The methodological quality and risk of bias of all included studies were assessed using the Prediction model Risk Of Bias Assessment Tool (PROBAST, version 1.0), which is specifically designed for evaluating studies that develop, validate, or update predictive models (de Jong et al., 2021). PROBAST was chosen because machine learning applications in talent identification constitute predictive modeling studies, and the tool allows systematic evaluation across relevant domains. To complement this formal appraisal, we also considered a broader construct of practical trustworthiness - the extent to which a model’s reported performance can be reasonably trusted for real-world decision support. This concept integrates three key safeguards: (i) external validation on independent data to test generalizability; (ii) calibration assessment to ensure probabilistic predictions correspond to observed outcomes; and (iii) data-leakage control, referring to methodological steps that prevent overlap between training and test information.

The PROBAST framework consists of four domains (Wolff et al., 2019): (i) participants, assessing whether the study sample is representative and appropriate for the intended target population; (ii) predictors, evaluating the definition, measurement, and availability of input variables; (iii) outcomes, assessing whether outcome definitions, timing, and measurement were appropriate; and (iv) analysis, focusing on modeling methods, handling of overfitting, missing data, validation, and performance reporting. Each domain includes signaling questions that guide judgments of "low," "high," or "unclear" risk of bias. An overall risk of bias judgment was made for each study by aggregating across domains, with studies classified as "low risk" only if all domains were rated low. If one or more domains were judged as high risk, the overall classification was high; if one or more were unclear with none rated high, the overall classification was unclear.

Two reviewers independently performed the risk of bias assessment for each included study. Discrepancies in judgments were resolved through discussion. All judgments were based exclusively on information reported in the published articles.

Given the particularities of machine learning research, special attention was given to signaling questions within the analysis domain, including handling of class imbalance, prevention of data leakage, adequacy of validation strategies, and transparency of reporting model performance metrics.

Effect measures

For the outcome domain - predictive performance of machine learning models for talent identification in team sports - we extracted and reported all performance metrics provided by the original studies. Given the diversity of machine learning methods and outcome definitions, no single effect measure was imposed a priori. Instead, the following effect measures were prioritized based on their frequency of use and interpretability in predictive modeling research.

For binary classification outcomes (e.g., selected vs. deselected, progressed vs. not progressed), the principal effect measures were overall accuracy, sensitivity (recall), specificity, precision (positive predictive value), F1-score, and the area under the receiver operating characteristic curve (AUC-ROC). Where reported, the area under the precision–recall curve (AUC-PR) was also extracted to account for class imbalance, which is common in talent identification contexts. For multi-class or clustering outcomes (e.g., player profiles, position-specific categories), measures such as overall classification accuracy, macro- and micro-averaged F1-scores, and adjusted Rand index were extracted.

For continuous outcomes (e.g., predictive regression of performance scores or advancement probabilities), effect measures included mean absolute error (MAE), root mean square error (RMSE), and coefficient of determination (R^2). Where multiple metrics were presented for the same model, all were recorded, but in synthesis greater emphasis was placed on metrics reflecting generalizability, particularly those derived from independent test sets or external validation cohorts.

No thresholds for minimally important differences were defined a priori, as such benchmarks do not currently exist for talent identification in team sports. Instead, results were interpreted with reference to established conventions in machine learning research (e.g., AUC-ROC values of 0.50 indicating no discrimination, 0.70 - 0.80 acceptable, 0.80 - 0.90 excellent, and >0.90 outstanding performance) while acknowledging the limitations of applying generic thresholds to heterogeneous sporting contexts.

No re-expression of results into alternative effect measures was required, as extracted metrics were analyzed in their originally reported form. The choice to retain multiple performance measures was justified by the heterogeneous reporting practices in the field and by the need to provide a transparent overview of predictive model performance rather than privileging a single effect measure.

Synthesis methods

Data from included studies were extracted into structured evidence tables designed to enable consistent cross-study comparison. Extraction focused on: (i) study identification details (sport, competitive level, and sample characteristics); (ii) input data domains (e.g., anthropometric, physical, technical, perceptual–cognitive, psychosocial, or multi-domain); (iii) machine learning approach (e.g., supervised classification, regression, ensemble learning, clustering, or deep learning methods); (iv) type of outcome predicted (e.g., selection vs. deselection, progression, positional classification, performance prediction, profiling, or maturation); (v) validation strategy and performance metrics; (vi) interpretability analyses or insights reported by authors; and (vii) main results and conclusions.

If studies tested multiple algorithms, results were extracted for each model, though synthesis tables emphasized the best-performing or most interpretable approach. No data transformations, imputations, or re-analyses were performed; where performance metrics or validation details were missing, these were reported as "not reported."

To facilitate synthesis, studies were grouped according to their primary analytic aim rather than by sport or algorithm. Each table followed a standardized column structure (General Aim, Outcomes Predicted, Key Performance Metrics, Interpretability/Key Insights, and Main Results & Conclusions). To improve clarity, abbreviation glossaries were provided for each table, and narrative overviews were written to introduce and contextualize the included studies.

Given the heterogeneity of sports, data modalities, machine learning methods, and outcome definitions, statistical pooling or meta-analysis was not feasible. Instead, a structured narrative synthesis was undertaken. This narrative integrated the tabular evidence with cross-cutting themes, focusing on: (i) recurring methodological patterns; (ii) relative strengths and limitations of different ML approaches; (iii) the role of interpretability in practical application; and (iv) conceptual insights into how ML has been used in talent identification and development.

Results

Study selection

A total of 228 records were identified through database searches (PubMed, n = 28; Scopus, n = 128; Web of

Science, n = 72). After removal of 83 duplicates, 145 records were screened by title and abstract, of which 63 were excluded. The remaining 82 reports were retrieved in full text, with none unretrievable. Following detailed eligibility assessment, 55 reports were excluded, primarily due to population not meeting inclusion criteria (n = 53) or intervention/outcomes not relevant (n = 2). Ultimately, 27 studies fulfilled all criteria and were included in the systematic review (Figure 1).

Study characteristics

Across the 27 studies included in this review, most (n=13) focused exclusively on football (soccer), reflecting its global prominence in youth talent pathways (Zhao et al., 2019; Jauhiainen et al., 2019; Abidin, 2021; Owen et al., 2022; Kelly et al., 2022; Abidin and Erdem, 2025). Other team sports examined included Australian Rules Football (Woods et al., 2018b; a; Gogos et al., 2020; Jennings et al., 2024), rugby (Woods et al., 2018a; Owen et al., 2022), basketball (Ge, 2024; Contreras-García et al., 2024), cricket (Brown et al., 2024), and volleyball (Formenti et al., 2022; Sanjaykumar et al., 2024). Sample sizes varied considerably, from very small academy samples as n=21 (Abidin, 2021) or n=22 (Formenti et al., 2022) to large federated datasets as 13,876 (Altmann et al., 2024) or n=2222 (Abidin and Erdem, 2025). While most studies reported male-only samples, some included both sexes (de Almeida-Neto et al., 2023; Ge, 2024) or were female-focused (Formenti et al., 2022; Sanjaykumar et al., 2024). Reporting of biological maturation was inconsistente, since some reported explicitly the maturation (de Almeida-Neto et al., 2023; Brown et al., 2024; Duncan et al., 2024), others not reported in many academy datasets (Altmann et al., 2024; Abidin and Erdem, 2025) (Table 1).

In terms of data domains, studies frequently combined anthropometric and physical performance measures (Craig and Swinton, 2021; de Almeida-Neto et al., 2023; Ge, 2024), but increasingly incorporated technical, psychological, perceptual–cognitive, or socio-cultural variables (Owen et al., 2022; Formenti et al., 2022; Brown et al., 2024). Supervised ML approaches predominated, with common algorithms including Random Forest (Abidin, 2021; Owen et al., 2022), Support Vector Machines (Razali et al., 2017; Abidin, 2021), penalized regression (Craig and Swinton, 2021; Kelly et al., 2022), and neural networks (de Almeida-Neto et al., 2023; Jennings et al., 2024). A smaller subset used unsupervised or hybrid approaches for clustering or anomaly detection (Jauhiainen et al., 2019; Ge, 2024; Contreras-García et al., 2024). Validation practices varied: while some employed robust strategies such as nested cross-validation (Altmann et al., 2024) or prospective external testing (Jennings et al., 2024), others relied only on internal resampling or leave-one-out (Razali et al., 2017; Formenti et al., 2022). Reporting of interpretability methods was uneven since some studies (Retzepis et al., 2024; Altmann et al., 2024) applied SHAP values, while others (Woods et al., 2018b; Abidin, 2021) relied on simpler feature rankings, and many did not address interpretability at all (Theagarajan and Bhanu, 2021; Sanjaykumar et al., 2024).

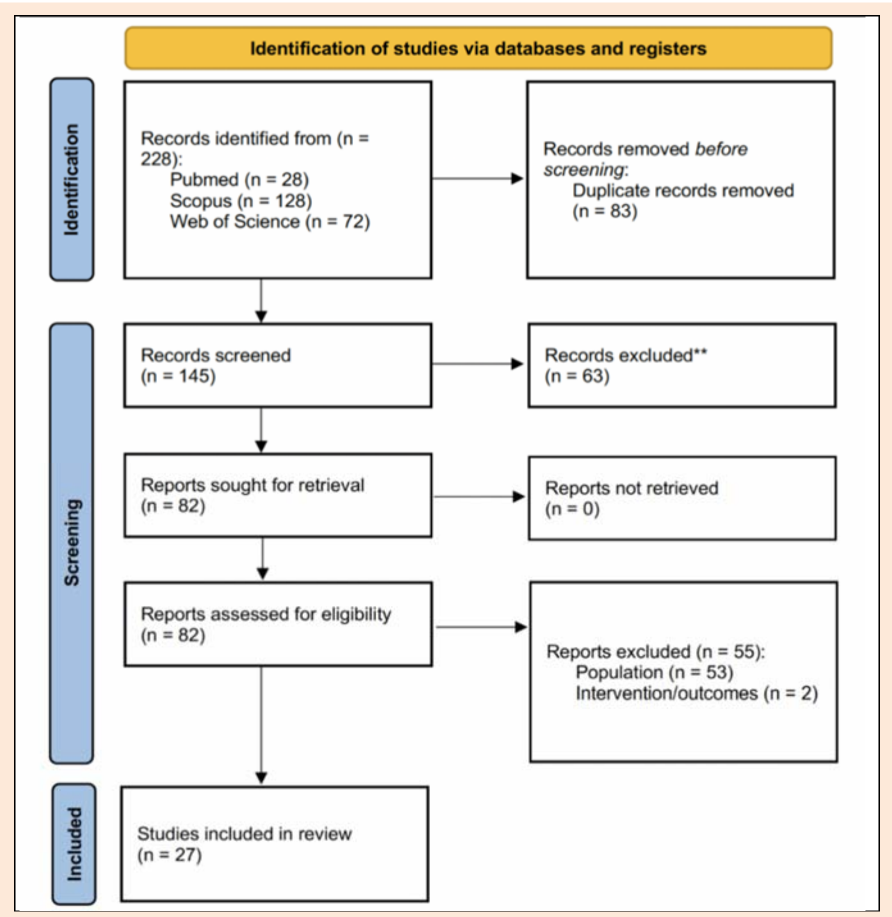

Figure 1. **PRISMA flow diagram (Page et al., 2021a).**

Table 1. Characteristics of the included studies.

Study	Sport	Competitive Context	Sample (n, sex, age range, maturation)	Data Domains & Sources	ML Approach (class, algorithms, model development)	Validation Strategy
(Abidin and Erdem, 2025)	Multiple (Football, Basketball, Volleyball, Athletics; plus "Others")	Sports high school entrance (youth selection, ages 14–16)	n=2222 (620Females/1602Males), 14–16 y, maturation not reported	Device-based physical tests (coordination via Spark, 30m sprint, vertical jump via JumpR, rhythm test) + coach evaluations (17 criteria: physical, reaction, specialism, psychological)	Deep learning (Shallow Deep Learning for Stage 1; novel Split-Combine-Merge Deep Learning [SCM-DL] for Stage 2); compared with Random Forest, Decision Tree, Extra Trees, SVC; nine feature selection methods (RFE variants, SelectKBest, Lasso, Boruta)	Train/test splits (70/30 and 80/20), k-fold cross-validation (3, 5, 7 folds); comparisons with multiple classifiers
(Abidin, 2021)	Football (soccer)	Altınordu Football Academy, U13 youth	n=21 field players (goalkeepers excluded), all male, age ≈ 13 y, maturation not reported; synthetic augmentation expanded to 231 instances	Training performance data via Hit/it Assistant (reaction times, coordination, speed, agility, etc.) + Coach evaluations across 18 qualitative/quantitative criteria (converted to numeric)	Supervised learning; seven algorithms tested in WEKA: ANN (MLP), SVM (SMO), Logistic Model Tree (LMT), Logistic regression, Naïve Bayes, Random Forest, CART. Dataset combined real + synthetic instances; preprocessing included normalization and derived position scores (D/M/F)	10-fold cross-validation
(de Almeida-Neto et al., 2023)	Football (soccer)	National-level youth athletes (club teams, ~V-level competition) and sports initiation program	n=75 males, 12–16 y (mean 13.3 ± 1.65), ~13% SI practitioners, 87% athletes; somatic maturation estimated (PHV categories: pre-, circa-, post-PHV)	Morphological (anthropometry, DXA: body mass, height, leg length, sitting height, body composition, BMC/BMD) + Neuromuscular (handgrip, medicine ball throw, vertical jump, countermovement jump via force plate)	Supervised deep learning; multilayer perceptron (MLP) artificial neural networks with backpropagation; z-scores used to normalize by sport/age; tested morphological, neuromuscular, and combined models	Train/test split (70/30) with cross-validation (10 repeated runs; all participants rotated through training/testing); ~10,000 training iterations
(Altmann et al., 2024)	Football (soccer)	German Bundesliga youth academy (U12–U19)	n=13,876 players (96% male), 11–19 y; maturation not explicitly reported but age categories (U12–U19) considered	Longitudinal match-derived data: ~32 million events across 10 years; position-specific technical/tactical features; aggregated spatiotemporal event-based data	Supervised ML; Gradient Boosted Decision Trees (LightGBM); models built separately per playing position; hyperparameter tuning with Bayesian optimization; features reduced with domain knowledge + automated selection	Nested cross-validation (inner loop for hyperparameter optimization, outer loop for model evaluation); train/test splits by season; temporal separation to avoid leakage
(Brown et al., 2024)	Cricket	County Age Group (CAG) programme, final trial stage	n=82 male players, 14–17 y (mean 15.3 ± 1.1); selected n=33, non-selected n=49; ethnicity: White British n=34, British South Asian n=44, Other n=4; maturation estimated (age at PHV, maturity offset)	Multidimensional: (a) physiological & anthropometrical (Yo-Yo test, sprint tests, jumps, planks, body size, weight, PHV), (b) perceptual–cognitive (video occlusion batting test), (c) psychological (PCDEQ + multiple psychosocial questionnaires), (d) participation history (practice/game history, multi-sport), (e) socio-cultural (ethnicity, schooling, siblings, birth quarter, postcode)	Supervised ML: Bayesian binomial regression (rSTAN); dimensionality reduction via correlation clustering → 21 derived features; weak normal prior	Cross-validation not reported; model convergence checks (posterior intervals, n_eff, BFMI) used for validation; sensitivity to ethnicity effects tested with interaction models
(Contreras-García et al., 2024)	Basketball	Spanish U14 Minicopa (youth) vs. Liga Endesa (professionals, comparator group)	n=217 U14 male players, 13–14 y; n=391 professional players; maturation not reported	Match-derived shooting charts (field goal attempts by location, 2020–21 & 2021–22 seasons)	Unsupervised ML (k-means and KNN clustering to classify shooting zones); outlier detection (IQR-based model) to identify "specialist shooters"	5-fold cross-validation for cluster classification; train/test split (20/80) for KNN consistency

Table 1. Continue…

Study	Sport	Competitive Context	Sample (n, sex, age range, maturation)	Data Domains & Sources	ML Approach (class, algorithms, model development)	Validation Strategy
(Cornforth et al., 2015)	Australian Rules Football	Elite professional players (AFL)	n=44 males, mean age 20 y, ~85.7 kg; maturation not reported	Physiological: daily ECG-derived HRV measures (time-, frequency-, and non-linear domain); Contextual: field size dimensions, match-day temperatures; Performance outcomes: GPS-derived match load, distance, speed zones	Supervised ML regression; seven algorithms in WEKA: Gaussian Processes, Linear Regression, LeastMedSq, Multilayer Perceptron, PLS Classifier, RBF Network, SMOreg; feature selection via PCA vs. wrapper subset + Genetic Algorithm	10-fold cross-validation; train/holdout splits tested
(Craig and Swinton, 2021)	Football (soccer)	Elite Scottish soccer academy (U10–U17), 10-year follow-up	n=512 male players, aged 10–17 at entry; 100 awarded pro contracts; maturation not directly reported; strong relative age effect observed	Anthropometric (height, weight, BMI) and physical performance (5, 10, 20m sprint times; countermovement jump; Yo-Yo IR1) collected longitudinally (1–14 sessions/player)	Supervised ML: LASSO logistic regression (with mixed-effects models for associations); multiple imputation for missing data	10-fold cross-validation to tune LASSO penalty; boot-strap (10,000 samples); train/test split (2/3–1/3) for predictive evaluation
(Duncan et al., 2024)	Football (soccer)	Grassroots club football in England (County FA structure)	n=162 boys, 7–14 y (mean 10.5 ± 2.1); biological maturation via APHV (Moore equation)	Anthropometry; maturity offset (APHV); fundamental movement skills via TGMD-3 with video scoring; perceived physical competence (PPASC); physical fitness (15 m sprint speed—timing gates; standing long jump); coach ratings (technical, social, physical, effort, overall); birth-quartile; technical skill test: UGent dribbling (procedural details reported)	Supervised ML: linear, ridge, lasso, random forest, boosted trees; recursive feature elimination; L1/L2 regularisation; collinearity control; Python implementation	Train/validation/test split 80/10/10 per age band; 5-fold cross-validation; age-band stratification to avoid leakage/under-representation
(Formenti et al., 2022)	Volleyball	Youth Italian championship, regional vs. provincial levels	n=26 female players (13 regional, 13 provin-cial), 13–15 y; maturation not reported	Volleyball-specific skill battery (setting, passing, spiking, serving; accuracy + technique); Physical performance (modified T-test COD, CMJ); Cog-nitive (Flanker task – executive control; Visual Search task – perceptual speed)	Supervised ML: Linear Discriminant Analysis, Logistic Regression, SVM, Decision Tree; features = volleyball skills + physical + cognitive measures	Stratified 5-fold cross-vali-dation
(Ge, 2024)	Basketball	Secondary school training teams	n=40 (20 boys, 20 girls), adolescents (~13–15 y); maturation not reported	Physical fitness tests (lung capacity, standing long jump, grip strength, 1000 m run boys / 800 m run girls); ~5000 test data entries used for model training/validation	Unsupervised feature learning (CNN + Autoencoder); Gaussian Mixture Model (EM algorithm for parameter estimation); model termed CNN-AE-MG	Train/test split 4:1 (4000/1000 records); ablation comparisons vs. CNN, CNN-AE, CNN-AE-SG; consistency tested with Bland-Altman plots
(Gogos et al., 2020)	Australian Rules Football	AFL U18 National/State/other combines; relates combine to senior career outcomes (retired/delisted cohort)	n=1,488 combine attendees (1999–2016); summary models on n=536 with ≥1 AFL player rating; mean age ≈18.5 y	Combine anthropometrics & physical tests (e.g., 20 m sprint, Yo-Yo IR, jumps), plus draft order & position; career outcomes from AFLTables & Champion Data	Linear models for ratings/ rankings; boosted regression trees for matches played (gradient-boosted ML)	Model fit assessed with BIC; no external validation; retrospective explanatory analysis

Table 1. Continue…

Study	Sport	Competitive Context	Sample (n, sex, age range, maturation)	Data Domains & Sources	ML Approach (class, algorithms, model development)	Validation Strategy
(Jauhiainen et al., 2019)	Football (soccer)	National TID database; focus on 14-y Finnish juniors and "academy player" labelling	N=951 14-year-old boys; minority "academy" class n≈14; tests/events 2011–2017	Physical tests (technical, speed, agility) + self-assessment (perceived competence, tactical skills, motivation) collected at biannual events; several data representations (phys, quest, combined)	One-class SVM (RBF) framed as anomaly detection to flag potential elite; PCA for decorrelation; k-NN imputation	Performance evaluated with AUC-ROC on held labels after unsupervised training; mean AUC ~0.763 across hyperparameters
(Jennings et al., 2024)	Australian Rules Football	Elite-junior AFL talent pathway; prospective prediction of 2021 National Draft	n=708 males; train 2017–2020 (n=465), prospective test 2021 (n=243)	Physical testing, in-game movement (GPS), and technical involvements; league-wide multi-season dataset	Logistic regression vs neural networks to classify drafted vs not drafted; operating at multiple cut-off thresholds	Prospective external hold-out (2017–20→2021) with sensitivity/specificity/ accuracy comparisons
(Kelly et al., 2022)	Football (soccer)	English professional academy; U9–U16 development and U18 selection/ deselection (contract)	Study 1: n=98, U9–U16; Study 2: n=18, U18 (male)	Multidomain 53 features across 8 methods over 2 seasons (technical/tactical, physical, psychological, social; e.g., PCDEQ, maturation %PAH, match hours)	Penalized regression (cross-validated LASSO via glmnet) predicting (a) review ratings; (b) achieving a pro contract	Cross-validation (CV) noted for LASSO; internal only
(Kilian et al., 2023)	Football (soccer)	Youth elite soccer talent-promotion program (DFB) — methodological evaluation on real program data	Sample details not fully specified in abstract text; applied to a set of multidimensional performance assessments within the program (youth cohort)	Multidomain performance battery used for latent factor structure; evaluation contrasts with PCA; study funded by DFB talent program	Deep latent-variable factor model: VAE estimator with importance-weighted variational inference + normalizing-flow priors; linear, identifiable measurement model (generalized EFA)	Robustness discussed; no classic predictive CV—focus is dimensionality reduction and identifiability; (not a selection classifier)
(López-De-Armentia, 2024)	Football (soccer)	Multi-league women's scouting context; data scarcity/coverage issues addressed by tool	~12,000 players tracked across ~30 leagues; basic roster & participation info (adults and youth)	Aggregated web-sourced player metadata (age, position, height, market value, contracts, injuries) and minutes played; alert generation pipeline	Rule-/criteria-driven alerts; "AI-powered" extraction mentioned, but no supervised model for TID classification is specified	Expert usability evaluation; no predictive CV/hold-out
(Owen et al., 2022)	Rugby Union	Regional age-grade academy selection (U16 & U18) in North Wales; talent camps	n=104 male; Mage=15.47±0.80; U16 n=62; U18 n=42; 66 selected/38 not	21 physiological (demographics, anthropometrics, sprint/power, grip, etc.) + 47 psychosocial (burnout, motivation, trait measures, EI, coping) assessed at selection days	Bayesian pattern-recognition pipeline to classify selected vs non-selected; position-specific models (forwards/backs)	Leave-one-out cross-validation (LOOCV) to minimize overfitting; internal validation only
(Razali et al., 2017)	Football (soccer)	Bukit Jalil Sports School (academy)	n=100; 15–17 y; sex not reported	Coach-rated physical, mental, and technical skills (1–10); Football Player Information System (BJSS)	Supervised classification; Bayesian Networks, Decision Tree, k-NN; WEKA implementation; GK excluded	Leave-one-out CV (small sample size)
(Retzepis et al., 2024)	Team sports	Preadolescent (≈11 y) team-sport athletes	n≈92; ~11 y; sex not reported	Anthropometry & motor tests (e.g., leg length, sitting height, weight, jumps) used to classify PHV	Supervised classification; Random Forest, Logistic Regression, Neural Network; forward feature selection with stratified 10-fold CV	10-fold stratified cross-validation (feature selection & tuning)
(Sandamal et al., 2024)	Football (soccer)	University-level players in Karakalpakstan vs. Khwarazm	n=60; 18–22 y; male	33 features (anthropometric, psychological, physical); questionnaires & field tests	Supervised regression/classification; Linear model, k-NN, Random Forest, XGBoost; SHAP for feature ranking	Train/test split with repeated evaluations; details limited

Table 1. Continue…

Study	Sport	Competitive Context	Sample (n, sex, age range, maturation)	Data Domains & Sources	ML Approach (class, algorithms, model development)	Validation Strategy
(Sanjaykumar et al., 2024)	Volleyball (women)	College-level players (state & national level)	n not reported; college-aged (≥18 y); female	Technical skill and execution metrics; likely field-based assessments	Supervised regression; KNN, Multiple Linear Regression, Lasso, Ridge, Elastic Net, Random Forest, XGBoost	Model evaluation via MAE, MSE, R²; split/CV details not reported
(Theagarajan and Bhanu, 2021)	Football (soccer)	High-school and professional competitions (video)	Image dataset: 49,950 images; includes high-school (youth) and pros	Match video frames; automated player/team/ball detection; event detection	Deep learning computer vision (object detection/tracking; event detection); supervised	Runtime and accuracy metrics discussed; formal CV/test split not reported
(Venkataraman et al., 2024)	Football (soccer)	Conceptual scouting framework; professional case studies	Sample not reported; case studies (e.g., Kevin De Bruyne); adults	Perceptual–cognitive attributes via YUVA-SQ questionnaire	None (scouting tool; no ML modeling)	Not applicable
(Woods et al., 2018b)	Australian Rules Football (AFL)	Elite junior (U18 national championships)	n=244 players; 680 observations; 17.6 ± 0.6 y; male	12 in-game technical skill indicators (match statistics)	Supervised classification; LDA, Random Forest, PART (decision list); variable importance & rule extraction	Internal classification accuracy; external validation not reported
(Woods et al., 2018a)	Rugby League	Elite youth (U20) vs. senior (not reportedL) competition comparison	U20: 372 obs; not reportedL: 378 obs; male	Team performance indicators from matches (not reportedL & U20)	Supervised classification tree to distinguish competitions; interpretable rules	Internal classification (apparent accuracy); external validation not reported
(Zhao et al., 2019)	Multi-sport (elite youth)	Elite sport school (6 sports; U15–U16)	n=97; male; U15–U16; training load ~20.8 h/week	18 anthropometric, 5 physiological, 2 motor tests; standardized lab/field assessments	Supervised multiclass classification; Linear Discriminant Analysis; Multilayer Perceptron; stepwise DA; repeated MLP training/testing	Leave-one-out (DA); repeated 80/10/10 splits for MLP

The Figure 2 summarizes the distribution of methodological rigor across different machine learning approaches used in youth-focused talent identification and development research. The chart highlights that most studies employed supervised, non-deep learning models with cross-validation as the primary evaluation method, while nested, temporal, or external validation approaches were rare.

Risk of bias in studies

Across the 27 included studies, the PROBAST assessment (Table 2) showed that 19 studies (70.4%) were rated Low risk of bias for Participants, and 20 studies (74.1%) for Predictors. In contrast, 13 studies (48.1%) were rated Unclear for Outcomes, and 13 studies (48.1%) were judged High risk in Analysis. Overall, 16 studies (59.3%) were assessed as having High risk of bias. Regarding applicability, 11 studies (40.7%) raised Some concern for Participants, 15 studies (55.6%) were rated Low concern for Predictors, and 11 studies (40.7%) were judged as having High concern for Outcomes.

Synthesis of studies

Table 3 synthesizes studies that focus primarily on selection prediction within talent identification systems, where ML models were used to determine whether athletes would be admitted, retained, or promoted at different stages of development. These works investigated diverse sports and settings, ranging from youth soccer academies (Jauhiainen et al., 2019; Craig and Swinton, 2021; Altmann et al., 2024), cricket county programes (Brown et al., 2024), rugby union regional selection (Owen et al., 2022), and Australian football drafts (Jennings et al., 2024). Studies also included models for admission and branch allocation in sport schools (Abidin and Erdem, 2025), as well as selection support tools for school athletes (Theagarajan and Bhanu, 2021).

Table 4 summarizes studies that applied ML to predict technical or physiological performance outcomes in sport. A study (Cornforth et al., 2015) reealed that regression models using pre-match heart rate variability (HRV) and environmental data could predict in-game outputs in Australian football. More recent studies employed ML to model skill-specific outcomes in youth soccer, such as dribbling performance (Duncan et al., 2024) and test-based fitness under environmental stressors (Sandamal et al., 2024). Similarly, Sanjaykumar et al. (Sanjaykumar et al., 2024) showed that Random Forest and XGBoost could accurately predict volleyball performance from anthropometric and body composition data.

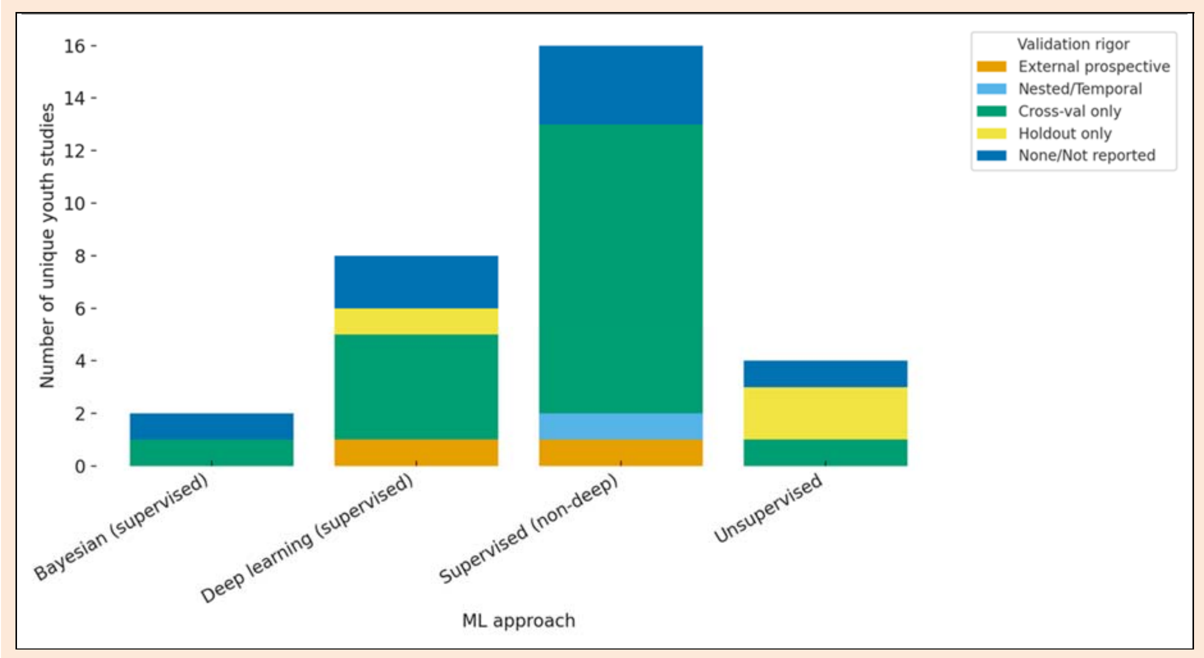

Figure 2. **Validation rigor and machine learning approaches in youth talent identification studies.**

Table 2. **Risk of bias assessment using PROBAST.**

Study	PROBAST Participants	PROBAST Predictors	PROBAST Outcomes	PROBAST Analysis	Overall ROB	Applicability (Participants)	Applicability (Predictors)	Applicability (Outcomes)
(Abidin and Erdem, 2025)	Low	Unclear	Low	High	High	Some concern	Some concern	Low concern
(Abidin, 2021)	Low	Unclear	Low	High	High	Some concern	Some concern	Low concern
(de Almeida-Neto et al., 2023)	Low	Low	High	Unclear	High	Some concern	Some concern	High concern
(Altmann et al., 2024)	Low	Low	Low	Unclear	Unclear	Some concern	Low–moderate concern	Low concern
(Brown et al., 2024)	Low	Low	Low	High	High	Some concern	Some concern	Low concern
(Contreras-García et al., 2024)	Unclear	Low	Unclear	Unclear	High	Low concern	Low concern	High concern
(Cornforth et al., 2015)	Low	Low	Unclear	Unclear	High	High concern	Low concern	High concern
(Craig and Swinton, 2021)	Low	Low	Low	Low	Low	Some concern	Low concern	Low concern
(Duncan et al., 2024)	Low	Low	Unclear	Unclear	Unclear	Some concern	Low concern	High concern
(Formenti et al., 2022)	Low	Low	Low	High	High	Some concern	Low concern	Moderate concern
(Ge, 2024)	Unclear	Unclear	Unclear	High	High	Unclear	High concern	High concern

Table 2. Continue…

Study	PROBAST Participants	PROBAST Predictors	PROBAST Outcomes	PROBAST Analysis	Overall ROB	Applicability (Participants)	Applicability (Predictors)	Applicability (Outcomes)
(Gogos et al., 2020)	Low	Low	Unclear	Unclear	Unclear	High concern	Low concern	High concern
(Jauhiainen et al., 2019)	Low	Low	Unclear	Unclear	Low	Low	Low	Low–moderate concern
(Jennings et al., 2024)	Low	Low	Low	Low/Unclear	Low	Low concern	Low concern.	Low concern
(Kelly et al., 2022)	Low	Low	Unclear	Unclear	Low	Low	Low	Low–moderate concern
(Kilian et al., 2023)	Low	Low	High	High	High	Low concern	Low concern	High concern
(López-De-Armentia, 2024)	Low	Some concerns	High	High	High	Some concern	High concern	High concern
(Owen et al., 2022)	Low	Low	Unclear	Unclear	Low	Low	Low	Low–moderate concern
(Razali et al., 2017)	Unclear	Low	High	High	High	Unclear	Low	Low–moderate concern
(Retzepis et al., 2024)	Low	Low	Low	Low	Low	Low concern	Low concern	Low concern
(Sandamal et al., 2024)	Unclear	Low	Unclear	High	High	Some concern	Low	High
(Sanjaykumar et al., 2024)	Unclear	Unclear	High	High	High	Unclear	Unclear	High concern
(Theagarajan and Bhanu, 2021)	Unclear	Low	High	High	High	Some concern	Some concern	High
(Venkataraman et al., 2024)	Unclear	Unclear	Unclear	High	High	Unclear	Unclear	Unclear
(Woods et al., 2018b)	Low	Low	Unclear	Unclear	Low	Low	Low	Low–moderate concern
(Woods et al., 2018a)	Low	Low	Unclear	Unclear	Low	Low	Unclear	Low–moderate concern
(Zhao et al., 2019)	Unclear	Unclear	Unclear	High	High	Unclear	Unclear	Unclear

Table 3. Synthesis of individual studies focusing exclusively in selection prediction.

Study	General Aim	Outcomes Predicted	Key Performance Metrics	Interpretability / Key Insights	Main Results & Conclusions
(Abidin and Erdem, 2025)	Selection Prediction	Stage 1: Admission (Pass/Fail). Stage 2: Branch allocation (Football, Basketball, Volleyball, Athletics, Other).	Stage 1: 98.9% accuracy (SDL). Stage 2: 97.4% accuracy, MCC 96.6% (SCM-DL, 6 features).	Feature selection revealed 6 key features spanning device tests & coach ratings; novel SCM-DL architecture captured hierarchical relations.	Authors conclude SCM-DL outperforms classical ML, can generalize to hierarchical datasets, and helps coaches prioritize features. External validity remains untested.
(Altmann et al., 2024)	Selection Prediction	Selection vs. deselection to the next age group (U12–U19) in elite German youth soccer academy across 7 years.	Best model XGBoost: ROC-AUC 0.69, F1-score 0.84. Models more sensitive to "selected" than "deselected."	Physical & physiological factors (linear sprint, COD sprint, CMJ, aerobic speed reserve) and soccer-specific skill most influential. Psychological measures of medium importance; health, age, and position-related variables inconsistent.	Authors conclude physical and skill-related measures are most decisive in selection/deselection; psychological factors moderate contributors. Suggests focusing academy monitoring on speed, power, endurance, and soccer-specific skill. Limitations: internal validation only, moderate discriminative ability (AUC <0.70).

AUC = Area Under the Curve; AUC-PR = Area Under the Precision–Recall Curve; BMI = Body Mass Index; BSA = British South Asian; CAG = County Age Group; CMJ = Countermovement Jump; COD = Change of Direction; DT = Decision Tree; F1 = F1-score (harmonic mean of precision and recall); IR1 (YoYo IR1) = Yo-Yo Intermittent Recovery Test, Level 1; KNN = K-Nearest Neighbors; LD/LDA = Linear Discriminant Analysis; LOOCV = Leave-One-Out Cross-Validation; LR = Logistic Regression; MCC = Matthews Correlation Coefficient; NN = Neural Network; Q1–Q4 = Birth quartiles (Relative Age Effect); RAE = Relative Age Effect; RF = Random Forest; ROC-AUC = Receiver Operating Characteristic – Area Under the Curve; SCM-DL = Split–Combine–Merge Deep Learning; SDL = Shallow Deep Learning; SVM = Support Vector Machine.

Table 3. Continue…

Study	General Aim	Outcomes Predicted	Key Performance Metrics	Interpretability / Key Insights	Main Results & Conclusions
(Brown et al., 2024)	Selection Prediction & Profiling	Differences between selected vs. non-selected youth male cricketers (U14–17) and between White British (WB) vs. British South Asian (BSA) selected players in County Age Group (CAG) programmes.	Not accuracy-based: model estimated probability shifts. Positive predictors of selection: athleticism, wellbeing/cohesion, birth in Q2–Q3, older brothers. Negative predictors: higher psych. scores, antisocial behaviour, younger brothers/older sisters. Ethnic group differences observed in athleticism, wellbeing, distress, antisocial behaviour.	Multidimensional input: 104 characteristics across 5 domains (physiological, perceptual-cognitive, psychological, participation history, socio-cultural). Analysis identified interaction between family structure, socio-cultural factors, and selection outcomes.	Authors conclude both athletic and socio-cultural variables play significant roles in selection. Highlight disparities: despite high BSA participation in grassroots cricket, nder-representation persists at selection level. Suggest systemic bias may influence CAG selection. Findings exploratory; sample small (N=82).
(Craig and Swinton, 2021)	Selection Prediction	Whether anthropometric (height, mass, BMI) and physical performance tests (20m sprint, CMJ, YoYo IR1) predict awarding of professional contracts in an elite Scottish soccer academy over 10 years.	Despite significant mean differences (successful players taller, faster, higher CMJ), predictive accuracy was near random: error proportion 0.43 (train), 0.45 (test) vs. 0.50 for random guessing.	Relative age effect (RAE) very strong: 50% of successful contracts born in Q1. CMJ, stature, and sprint had small associations but high overlap with non-successful players. No reliable case-level prediction possible.	Authors conclude that anthropometric and physical performance profiling alone cannot predict professional contract success within already talented academy players. Recommend data be used to guide training, not selection. Suggest holistic models integrating technical, tactical, psychological, and sociocultural variables, plus coach expertise. Stress need for addressing RAE bias (e.g., bio-banding, scout education).
(Formenti et al., 2022)	Selection Prediction	Classification of female junior volleyball players as regional vs. provincial level based on volleyball-specific skills, physical performance, and cognitive functions.	Decision Tree: Precision 93%, Recall 73%, F1 = 0.83. Other models (LD, LR, SVM) performed lower (Precision 47–63%, Recall 57–73%).	DT identified passing and spiking technique plus cognitive task response times (Flanker congruent/incongruent, Visual search 10/15 items) as key discriminators. Physical tests (COD, CMJ) contributed less.	Authors conclude that higher-level players outperform lower-level peers across volleyball skills, COD, CMJ, and cognitive functions. ML results emphasize the role of cognitive functions + technical skills (passing, spiking) in discriminating competitive level. Practical recommendation: include training of both volleyball-specific techniques and executive/perceptual skills in youth development.
(Jauhiainen et al., 2019)	Selection Prediction	Detection of potential elite youth soccer players (academy contracts) from large dataset of junior players (N=951, age 14).	Best performance with "phys large" dataset (N=951, 16 physical test variables): AUC-ROC = 0.763 (±0.007), AUC-PR = 0.960, Sensitivity = 0.80, Specificity = 0.61. Smaller sets ("phys+quest", "quest") performed worse (AUC-ROC 0.58–0.66).	Demonstrated utility of anomaly detection for imbalanced TID problems (14 academy vs. 937 non-academy). Physical tests (jump, sprint, agility) more predictive than questionnaire/self-assessment. Nonlinear SVM outperformed linear baseline.	Authors conclude that one-class SVM can moderately identify future academy players but specificity remains limited (many false positives). Results promising but not sufficient for stand-alone selection. Recommend larger datasets, longitudinal validation, and integration of multidimensional variables.

AUC = Area Under the Curve; AUC-PR = Area Under the Precision–Recall Curve; BMI = Body Mass Index; BSA = British South Asian; CAG = County Age Group; CMJ = Countermovement Jump; COD = Change of Direction; DT = Decision Tree; F1 = F1-score (harmonic mean of precision and recall); IR1 (YoYo IR1) = Yo-Yo Intermittent Recovery Test, Level 1; KNN = K-Nearest Neighbors; LD/LDA = Linear Discriminant Analysis; LOOCV = Leave-One-Out Cross-Validation; LR = Logistic Regression; MCC = Matthews Correlation Coefficient; NN = Neural Network; Q1–Q4 = Birth quartiles (Relative Age Effect); RAE = Relative Age Effect; RF = Random Forest; ROC-AUC = Receiver Operating Characteristic – Area Under the Curve; SCM-DL = Split–Combine–Merge Deep Learning; SDL = Shallow Deep Learning; SVM = Support Vector Machine.

Table 3. Continue…

Study	General Aim	Outcomes Predicted	Key Performance Metrics	Interpretability / Key Insights	Main Results & Conclusions
(Jennings et al., 2024)	Selection Prediction	Drafted vs. not-drafted players in the AFL National Draft (2021) using physical, GPS (in-game movement), and technical involvement data.	Neural networks consistently outperformed logistic regression: NN specificity = 79 ± 13%, sensitivity = 61 ± 24%, accuracy = 76 ± 8% vs. LR specificity = 73 ± 15%, sensitivity = 29 ± 14%, accuracy = 66 ± 11%. At draft-rate threshold (15%) and convergence threshold (35%), NN classified more drafted players in 88% of comparisons.	Neural networks handled unfactored, high-dimensional inputs better than LR, capturing nonlinear relationships. Logistic regression benefited only when data were factored (dimensionality reduction). Key insight: sensitivity (identifying drafted players) is paramount, and NN achieved superior balance of sensitivity and specificity.	Authors conclude that NN models are more effective than logistic regression for predicting draft outcome, particularly when identifying drafted players (sensitivity). Practical implications: clubs may apply NN-based models to complement subjective scouting and reduce bias. Limitations: data restricted to one state league, psychosocial variables absent, career success beyond draft not considered.
(Owen et al., 2022)	Selection Prediction	Selection vs. non-selection to regional U16 and U18 rugby squads based on 21 physiological and 47 psychosocial factors. Analyses run for all players, forwards, and backs.	Physiological models: 67.6% (all), 70.1% (forwards), 62.5% (backs). Psychosocial models: 62.3% (all), 73.7% (forwards), 60.4% (backs). Specificity higher than sensitivity in all cases.	Key physiological predictors: greater hand grip strength, faster 10m & 40m sprints, higher power and momentum. Key psychosocial predictors: lower burnout, reduced exhaustion, lower reduced sense of accomplishment, lower life stress (forwards), and lower difficulty describing feelings (forwards). For backs, lower interjected regulation and lower burnout were features.	Authors conclude physiological factors (strength, speed, power) are more predictive of rugby selection than psychosocial ones, but psychosocial variables (especially lower burnout and stress) also play a significant role. Position-specific differences exist (e.g., emotional regulation markers more relevant for forwards). Recommend holistic, position-tailored selection frameworks including psychosocial screening alongside physiological testing.
(Theagarajan and Bhanu, 2021)	Selection Support	Classification of students' sports-specific talent category (basketball, volleyball, football, athletics, kabaddi, weightlifting) based on anthropometric and physical fitness attributes.	Random Forest highest: 96.2% accuracy; SVM 95.5%; KNN 95.2%; Decision Tree 92.6%; Naïve Bayes 89.8%.	Feature importance analysis showed attributes like height, weight, speed, and endurance strongly influenced classification. Models could allocate students to most likely successful sport pathway.	Authors conclude ML, especially RF and SVM, can reliably classify school-level athletes into suitable sports, providing data-driven support for talent identification and allocation. Limitations: small, single-institution dataset; attributes mostly physical, excluding psychological/technical. Recommend broader variables and longitudinal validation.

AUC = Area Under the Curve; AUC-PR = Area Under the Precision–Recall Curve; BMI = Body Mass Index; BSA = British South Asian; CAG = County Age Group; CMJ = Countermovement Jump; COD = Change of Direction; DT = Decision Tree; F1 = F1-score (harmonic mean of precision and recall); IR1 (YoYo IR1) = Yo-Yo Intermittent Recovery Test, Level 1; KNN = K-Nearest Neighbors; LD/LDA = Linear Discriminant Analysis; LOOCV = Leave-One-Out Cross-Validation; LR = Logistic Regression; MCC = Matthews Correlation Coefficient; NN = Neural Network; Q1–Q4 = Birth quartiles (Relative Age Effect); RAE = Relative Age Effect; RF = Random Forest; ROC-AUC = Receiver Operating Characteristic – Area Under the Curve; SCM-DL = Split–Combine–Merge Deep Learning; SDL = Shallow Deep Learning; SVM = Support Vector Machine.

Table 4. Synthesis of individual studies focusing exclusively in performance prediction.

Study	General Aim	Outcomes Predicted	Key Performance Metrics	Interpretability / Key Insights	Main Results & Conclusions
(Cornforth et al., 2015)	Performance Prediction	Prediction of in-game performance in elite Australian football players using pre-match HRV measures (time, frequency, nonlinear domains) plus environmental/field data.	Best correlations with GA wrapper + regression algorithms: Walk r=0.76, Jog r=0.75, Cruise r=0.73, Player Load r=0.72, Match Distance r=0.73. PCA improved slightly over all-variables approach, but GA wrapper yielded the highest predictive performance (mean r=0.60 vs. 0.49–0.53).	Highlighted the value of advanced regression (esp. SMOreg, Gaussian Processes) combined with feature selection. Identified HRV-derived features (esp. nonlinear measures) plus environmental conditions (temperature, field size) as significant contributors to match performance.	Authors conclude sophisticated regression models can predict match performance >0.70 correlation from HRV and environmental data. Potential to support player selection decisions and training load adjustments tailored to field dimensions and match-day conditions. Early demonstration of sport informatics potential in team sport.
(Duncan et al., 2024)	Performance Prediction	Dribbling skill (UGent dribbling test, skill differential with/without ball).	Initial accuracy: linear ~57%, ridge ~48%, lasso ~34%, RF ~68%, boosted ~66%. When stratified by age band: RF 98.6%, boosted trees 96.1%, lasso 94.1%, linear 91.9%.	Feature importance: FMS score most influential, followed by coach overall rating, years of playing experience, and APHV. Birth quartile and chronological age least important.	ML showed technical skills can be predicted with high accuracy from multidimensional inputs, especially FMS. Supports theory that broad motor skill competence underpins technical soccer ability. Coaches should emphasize FMS training before sport-specific drills. Suggests a shift away from over-reliance on physical testing alone.
(Sandamal et al., 2024)	Performance Prediction	Prediction of soccer players' performance in field-based tests: Dribbling Shuttle Test (DSt), Goal Accuracy Test (GAt), and Yo-Yo Intermittent Recovery Test Level 1 (YYIRT1).	XGBoost consistently outperformed RF and KNN across tests (highest R^2 and lowest error). RF showed moderate accuracy, KNN lowest. Performance varied between cohorts, with Karakalpakstan athletes showing reduced predicted fitness values.	SHAP global explanations: anthropometric (sitting height, meso breadth), hematological, and hormonal markers (E2, IGF-1, cortisol, testosterone) emerged as top predictors. LIME local explanations confirmed hormonal differences: E2, IGF-1, cortisol strongly impacted fitness in environmentally exposed group, while testosterone was more influential in controls.	Authors conclude explainable ML (esp. XGBoost + SHAP/LIME) offers accurate and interpretable fitness prediction in young soccer players. Results highlight negative effects of environmental degradation (Aral Sea region) on hormonal balance and physical performance. Study demonstrates value of explainable AI for screening and tailoring training in vulnerable populations. Limitations: relatively small cohorts, region-specific findings, no external validation.
(Sanjaykumar et al., 2024)	Performance Prediction	Prediction of on-court performance based on demographic and physical attributes (age, height, weight, fat %, muscle mass, bone mass, BMI).	RF: R^2=0.9418, accuracy=94.18%, RMSE=2.67. XGBoost: R^2=0.9276, acc=92.76%, RMSE=2.98. Linear Regression weaker: R^2=0.7531, acc=75.31%, RMSE=5.51.	Correlation analysis: Height (r=0.879), muscle mass (r=0.653), bone mass (r=0.622) strongly positively related to performance. BMI not significant (r=0.04). RF captured nonlinearities best; XGBoost close.	Authors conclude ML—especially Random Forest—provides accurate and objective prediction of volleyball performance from physical attributes. Supports more data-driven talent ID, moving beyond subjective scouting. Future work: integrate skill and psychological factors, extend to diverse populations.

ACC = Accuracy; AUC = Area Under the Curve; APHV = Age at Peak Height Velocity; BMI = Body Mass Index; DSt = Dribbling Shuttle Test; FMS = Fundamental Movement Skills; GA = Genetic Algorithm; GAt = Goal Accuracy Test; HRV = Heart Rate Variability; IGF-1 = Insulin-like Growth Factor 1; KNN = K-Nearest Neighbors; LASSO = Least Absolute Shrinkage and Selection Operator; LIME = Local Interpretable Model-agnostic Explanations; PCA = Principal Component Analysis; R^2 = Coefficient of Determination; RF = Random Forest; RMSE = Root Mean Squared Error; SHAP = SHapley Additive exPlanations; SMOreg = Sequential Minimal Optimization regression; XGBoost = Extreme Gradient Boosting; YYIRT1 = Yo-Yo Intermittent Recovery Test, Level 1.

Table 5 compiles studies exploring the use of ML for team formation and playing position classification, where algorithms aim to replicate or optimize decisions traditionally made by coaches. A study (Abidin, 2021) tested multiple ML models for both position assignment and lineup generation in youth soccer, demonstrating high concordance with coach decisions. Other study (Razali et al., 2017) developed a prototype system to classify football players into positional roles using physical, mental, and technical ratings, validated by expert coach evaluation. Finally a study (Woods et al., 2018b) examined positional classification in elite junior Australian football using technical skill indicators, highlighting the limitations of conventional statistics for discriminating playing roles.

Table 6 includes studies that address broader or emerging applications of ML in talent identification and development, spanning orientation, specialization, profiling, maturation, and scouting support. Examples include orientation of youth into appropriate sports using morphological and neuromuscular profiles (de Almeida-Neto et al., 2023), detection of premature specialization in basketball (Contreras-García et al., 2024), fitness assessment with deep learning (Ge, 2024), and forecasting AFL career outcomes (Gogos et al., 2020). Other studies investigated multidimensional predictors of progression (Kelly et al., 2022), latent factor modeling of youth soccer assessments (Kilian et al., 2023), and scouting frameworks in women's and men's football (Venkataraman et al., 2024; López-De-Armentia, 2024). A study (Retzepis et al., 2024) applied explainable ML to maturation prediction, while other (Woods et al., 2018a) compared gameplay profiles of youth vs. senior rugby league. Finally a study (Zhao et al., 2019) demonstrated cross-sport profiling with anthropometric and physiological tests.

Table 5. Synthesis of individual studies focusing in playing position/team formation prediction.

Study	General Aim	Outcomes Predicted	Key Performance Metrics	Interpretability / Key Insights	Main Results & Conclusions
(Abidin, 2021)	Selection Prediction & Team Formation	Player position classification (Defender, Midfielder, Forward) and lineup formation for U13 Altınordu Football Academy players. Compared ML lineups with coach's ideal lineup and 20 match lineups.	RF best at 93.9% accuracy, κ=0.91; MLP 92.6%, LMT 90.5%. Adding Hit/it training data improved accuracy across all algorithms vs. baseline (e.g., RF 81.8% → 93.9%). For team formation, lineups of SMO & SimpleCART closest to coach (Pearson r≈0.975). Lineup similarity with match lineups averaged 89.36%.	Demonstrated importance of combining coach evaluation + training device (Hit/it) data. Synthetic data generation addressed small sample. Lineup similarity analysis showed ML can approximate coach/team decisions without using match data.	Authors conclude ML models (esp. RF, MLP, LMT) can reliably support player selection and lineup formation, potentially integrated into weekly coaching tools. Hit/it data deemed essential to boost predictive accuracy. External generalizability remains untested beyond single academy.
(Razali et al., 2017)	Selection Support & Team Formation	Prediction of most suitable playing position (10 outfield roles: sweeper, backs, midfielders, wingers, forwards; GK excluded) based on physical, mental, and technical ratings.	Bayesian Networks: 99% accuracy; Decision Tree: 98%; KNN: 97%.	Framework combined coach-rated attributes (1–10 scale across physical, mental, technical skills) with ML classifiers. Developed a Football Talent Identification Site for practical deployment. Expert evaluation (20 coaches/managers) confirmed ease of use and relevance.	Authors conclude ML classifiers can assign players to their optimal positions with very high accuracy, reducing subjective bias in coach decisions. Prototype system was well-received (75–80% strongly agreed on usability, suitability). Limitations: small single-school dataset, manual skill ratings subjective, no external validation.
(Woods et al., 2018b)	Team Formation & Position Classification	Classification of elite junior Australian football players (U18) into 4 playing positions (defender, forward, midfield, ruck) based on 12 technical skill indicators from national championships.	LDA: 56.8% accuracy (errors: midfielders 19.6% → rucks 75%). Random Forest: 51.6% accuracy (errors: midfielders 27.8% → rucks 100%). PART decision list: 70.1% accuracy (errors: midfielders 14.4% → rucks 100%).	Rule induction (PART) generated 6 classification rules, mainly leveraging disposals, contested/ uncontested possessions, kicks, and inside 50s. Showed defenders and forwards overlapped heavily; midfielders most distinct; rucks poorly classified due to small sample.	Authors conclude that existing commercial technical indicators provide limited discriminatory power for position classification, with high homogeneity across roles. PART offered relatively better accuracy but overfitting risk noted. Practical implication: recruiters should use more position-specific technical indicators and design competitions/training environments that allow players to demonstrate role-specific attributes. Reliance solely on standard technical stats may obscure positional differences and complicate objective recruitment.

GK = Goalkeeper; κ = Cohen's Kappa (agreement statistic); LDA = Linear Discriminant Analysis; LMT = Logistic Model Tree; MLP = Multilayer Perceptron; PART = Partial Decision List (rule-based classifier); RF = Random Forest; SMO = Sequential Minimal Optimization; SimpleCART = Classification and Regression Tree (simplified); U13/U18 = Under-13 / Under-18 age category.

Table 6. Main results of the individual studies on multiple objectives.

Study	General Aim	Outcomes Predicted	Key Performance Metrics	Interpretability / Key Insights	Main Results & Conclusions
(de Almeida-Neto et al., 2023)	Orientation & Selection Support	Predicted similarity between morphological + neuromuscular profiles of youth in Sport Initiation (SI) vs. young athletes in six sports (soccer, swimming, tennis, volleyball, rowing, BJJ).	Reliability of MLP models reported at 87%. Similarity scores: SI → Soccer 88%, Swimming 79%, BJJ 77%, Tennis 70% (combined analysis). No significant similarity for Rowing.	Demonstrated how MLPs can integrate morphological + neuromuscular + biological maturation factors. Highlighted BM as a major confounder influencing neuromuscular strength and morphology. Suggested that MLPs can reduce selection errors by combining multiple domains.	Authors conclude MLPs are effective tools to guide orientation of SI youth into sports matching their physical/neuromuscular profiles, reducing misallocation risk. Stress need to consider biological maturation in TID. Limitations: cross-sectional, small sample (N=75), no longitudinal follow-up, non-elite athletes.
(Contreras-García et al., 2024)	Development / Specialization Analysis	Classification of shooting zones and detection of outlier patterns to identify early specialization vs. versatility in U14 basketball players compared with professional players.	KNN model classification of shots reached 99.6% accuracy (professionals as reference). Outlier analysis: 97.7% of U14 players vs. 64.7% of professionals showed extreme FGA% patterns. Versatility: U14 2.3% vs. Professionals 35.4%.	Machine learning cluster analysis identified 8 shooting zones; combined with outlier detection, yielded 7 role categories. Revealed U14 lacked versatility and 3-point shooting ability, often over-specializing in 2–4 midrange zones. Professionals characterized by either versatile players or one-zone specialists.	Authors conclude U14 basketball players show premature specialization patterns not aligned with professional demands. Recommend formative training to enhance shooting versatility or to cultivate one-zone specialist roles deliberately. Findings raise concerns that current youth competitions may prioritize short-term success over long-term player development.
(Ge, 2024)	Performance Assessment & Training Support	Quantitative classification of youth basketball players' physical fitness (excellent, good, pass, fail) using CNN-AE-MG model.	CNN-AE-MG achieved mAP = 89.12%, assessment accuracy = 97.5%. Male subgroup prediction 100% accurate (20/20 correct), female subgroup 95% (19/20 correct).	Combination of CNN + Autoencoder enabled unsupervised feature learning, reducing feature loss. Gaussian Mixture with EM algorithm improved classification reliability. Identified endurance (1000m/800m), lung capacity, grip strength as weak areas in youth players.	Authors conclude the CNN-AE-MG model provides accurate, dynamic assessment of youth basketball players' physical health, superior to baseline models. Proposed use for exercise prescription personalization, training program adjustment, and talent selection support. Limitations: single-country, limited external validation, general fitness focus rather than sport-specific outcomes.

AE = Autoencoder; AFL = Australian Football League; APHV = Age at Peak Height Velocity; BJJ = Brazilian Jiu-Jitsu; BM = Body Mass; CNN = Convolutional Neural Network; CNN-AE-MG = Convolutional Neural Network – Autoencoder – Mixture Gaussian model; CI = Conditional Inference; DA = Discriminant Analysis; EM = Expectation–Maximization; F1 = F1-score (harmonic mean of precision and recall); FGA% = Field Goal Attempt Percentage; Hb = Hemoglobin; HR = Heart Rate; IGF-1 = Insulin-like Growth Factor 1; KNN = K-Nearest Neighbors; Lasso = Least Absolute Shrinkage and Selection Operator regression; LR = Logistic Regression; MLP = Multilayer Perceptron; nI-WAVE = Nonlinear Importance-Weighted Autoencoding Variational Inference with normalizing flow priors; NRL = National Rugby League; PCA = Principal Component Analysis; PCDEQ = Psychological Characteristics of Developing Excellence Questionnaire; PHV = Peak Height Velocity; RF = Random Forest; ROC AUC = Receiver Operating Characteristic – Area Under the Curve; SHAP = SHapley Additive exPlanations; SI = Sport Initiation; U20 = Under-20 age category; U12/U14/U18 = Under-12 / Under-14 / Under-18 age categories; YODA = Youth Online Diagnostic Assessment; YUVA-SQ = Youth Universal Value Assessment – Scouting Questionnaire.

Table 6. **Continue…**

Study	General Aim	Outcomes Predicted	Key Performance Metrics	Interpretability / Key Insights	Main Results & Conclusions
(Gogos et al., 2020)	Selection Prediction & Career Outcome Forecasting	Career outcomes of AFL draftees (matches played, mean AFL Player Rating, mean AFL Player Ranking).	Draft combine alone explained <3–4% of variance in career outcomes. Adding draft order & playing position improved variance explained slightly (up to 6%). Individual combine tests explained <2% variance.	Boosted trees showed player position (>35% relative importance) and draft order (>25%) far outweighed combine results (<10%). Key forwards showed no clear relation between draft position and in-game performance; midfielders/rucks showed positive relation. Evidence of loss aversion bias: early draftees played more games irrespective of performance.	Authors conclude AFL Draft Combine tests are poor predictors of long-term career outcomes. Draft position and playing position provide small additional explanatory power. Suggests physical test batteries are insufficient for TID and should be complemented by in-game skill, decision-making, and contextual factors. Highlights systemic biases (early draft order → more opportunities).
(Kelly et al., 2022)	Talent Development	(a) Player review ratings (U9–U16, n=98); (b) Selection to professional contract (U18, n=18). Both based on ~53 variables across four domains (technical/tactical, physical, psychological, social).	Study 1: 15/53 features had non-zero coefficients; strongest = % predicted adult height (0.196), lob pass (0.160), dribble completion (0.124), total match-play hours (0.145), older relative age. Study 2: strongest predictors of professional contract = PCDEQ Factor 3 (coping with pressures), PCDEQ Factor 4 (ability to organise quality practice), plus progression ratings, slalom dribble, lower home SES.	Lasso regression identified holistic, non-linear predictors across all FCM domains. Key insight: psychological factors (esp. coping with pressure, organization) emerged as strongest contributors to contract attainment, not just technical/physical. Also highlights relative age bias and importance of match-play opportunities.	Authors conclude that youth development is multifactorial and dynamic. Success not solely determined by technical/tactical ability; psychological resilience and self-organization are critical. Early maturation, relative age, and cumulative match-play also drive coaches' evaluations. Findings support bio-banding and greater investment in psychological development within academies. Limitations: small samples (esp. Study 2), retrospective data, exploratory nature of ML.
(Kilian et al., 2023)	Profiling / Latent Structure Analysis	Identification of latent factors underlying multidimensional assessments (technical, tactical, physical, anthropometric, psychosocial).	Not predictive classification; evaluated model fit and factor interpretability. nI-WAVE outperformed PCA with clearer separation, fewer cross-loadings.	Four interpretable latent factors: (1) Subjective coach evaluations, (2) Anthropometric/age-related (incl. sprint), (3) Technical skills (dribbling, ball control, juggling), (4) Speed/agility. nI-WAVE showed superior interpretability and factor structure stability.	Authors conclude that deep learning factor models (nI-WAVE) provide better latent structure recovery than PCA, improving interpretability of multidimensional TID data. Highlight importance of large-scale datasets in advancing ML-based profiling. Limitations: requires large data, anchors affect loadings, only U12 German cohort examined.

AE = Autoencoder; AFL = Australian Football League; APHV = Age at Peak Height Velocity; BJJ = Brazilian Jiu-Jitsu; BM = Body Mass; CNN = Convolutional Neural Network; CNN-AE-MG = Convolutional Neural Network – Autoencoder – Mixture Gaussian model; CI = Conditional Inference; DA = Discriminant Analysis; EM = Expectation–Maximization; F1 = F1-score (harmonic mean of precision and recall); FGA% = Field Goal Attempt Percentage; Hb = Hemoglobin; HR = Heart Rate; IGF-1 = Insulin-like Growth Factor 1; KNN = K-Nearest Neighbors; Lasso = Least Absolute Shrinkage and Selection Operator regression; LR = Logistic Regression; MLP = Multilayer Perceptron; nI-WAVE = Nonlinear Importance-Weighted Autoencoding Variational Inference with normalizing flow priors; NRL = National Rugby League; PCA = Principal Component Analysis; PCDEQ = Psychological Characteristics of Developing Excellence Questionnaire; PHV = Peak Height Velocity; RF = Random Forest; ROC AUC = Receiver Operating Characteristic – Area Under the Curve; SHAP = SHapley Additive exPlanations; SI = Sport Initiation; U20 = Under-20 age category; U12/U14/U18 = Under-12 / Under-14 / Under-18 age categories; YODA = Youth Online Diagnostic Assessment; YUVA-SQ = Youth Universal Value Assessment – Scouting Questionnaire.

Table 6. Continue…

Study	General Aim	Outcomes Predicted	Key Performance Metrics	Interpretability / Key Insights	Main Results & Conclusions
(López-De-Armentia, 2024)	Scouting Support & Talent Detection	Detection of potential women's football talents across ~30 leagues using automated data collection (Soccerdonna) + alert system.	No accuracy metrics (non-ML predictive model). Evaluation: Usefulness 4–5/5; Ease of use 4–5/5; all experts agreed alerts were effective and tool improved efficiency.	Tool integrates basic player data (demographics, position, minutes, contract expiry, market value, injuries) with automatic alert generation (e.g., U20 players with 1000 min, >5 goals, or consistent starts). Dashboards allow filtering/searching ~12,000 players.	Authors conclude WTDTool increases efficiency and coverage in scouting women's football, particularly for clubs with limited resources. Experts confirmed ease of use and usefulness. Limitations: women's data coverage incomplete (contract and market data available for only ~25% of players); no predictive analytics yet. Future: add anomaly detection and integrate multiple data sources.
(Retzepis et al., 2024)	Maturation Prediction	Classification of athletes with predicted PHV ≤ median vs. > median age, using anthropometric, body composition, and strength measures.	LR achieved 96.67% accuracy, 98% recall, 96.33% precision, 97.09% F1-score, ROC AUC 99%. RF and NN slightly lower (94–96%).	SHAP (explainable AI) revealed key predictors: sitting height, weight, height, body fat, left & right handgrip strength, father's height. Sitting height and weight most influential (higher values → PHV > median). Body fat higher predicted PHV ≤ median.	Study concludes explainable ML can accurately predict PHV timing in 11-year-old athletes. Key growth and strength indicators (esp. sitting height, weight, grip strength) discriminate maturity status. Findings help avoid misclassification of early maturers as "talents" and support better talent ID, injury prevention, and training load management. Recommends longitudinal validation to confirm predictive power and extend to other sports and female athletes.
(Venkata-raman et al., 2024)	Scouting Support & Cognitive Profiling	Player suitability for selection and development, integrating psychometric (YODA) and coach-based evaluations into a standardized scouting framework (YUVA-SQ).	Not accuracy-based: case demonstration. YODA generated trait/personality plots for individual players, producing actionable insights for coaches. Validated by expert use and player development outcomes.	YODA psychometric tool provided granular insights into players' cognitive profile (e.g., coachability, team orientation, game knowledge, analytical style). Combined with coach technical ratings and trial performance for continuous monitoring.	Authors conclude YUVA-SQ offers a holistic, standardized scouting framework blending cognitive/behavioral assessment with technical/physical evaluation. Demonstrated utility in restructuring a university football team. Proposed extension to grassroots talent scouting in India, aligning with AIFF "Vision 2047." Limitations: descriptive case study only, no predictive performance metrics, no large-scale validation.

AE = Autoencoder; AFL = Australian Football League; APHV = Age at Peak Height Velocity; BJJ = Brazilian Jiu-Jitsu; BM = Body Mass; CNN = Convolutional Neural Network; CNN-AE-MG = Convolutional Neural Network – Autoencoder – Mixture Gaussian model; CI = Conditional Inference; DA = Discriminant Analysis; EM = Expectation–Maximization; F1 = F1-score (harmonic mean of precision and recall); FGA% = Field Goal Attempt Percentage; Hb = Hemoglobin; HR = Heart Rate; IGF-1 = Insulin-like Growth Factor 1; KNN = K-Nearest Neighbors; Lasso = Least Absolute Shrinkage and Selection Operator regression; LR = Logistic Regression; MLP = Multilayer Perceptron; nI-WAVE = Nonlinear Importance-Weighted Autoencoding Variational Inference with normalizing flow priors; NRL = National Rugby League; PCA = Principal Component Analysis; PCDEQ = Psychological Characteristics of Developing Excellence Questionnaire; PHV = Peak Height Velocity; RF = Random Forest; ROC AUC = Receiver Operating Characteristic – Area Under the Curve; SHAP = SHapley Additive exPlanations; SI = Sport Initiation; U20 = Under-20 age category; U12/U14/U18 = Under-12 / Under-14 / Under-18 age categories; YODA = Youth Online Diagnostic Assessment; YUVA-SQ = Youth Universal Value Assessment – Scouting Questionnaire.

Table 6. Continue…

Study	General Aim	Outcomes Predicted	Key Performance Metrics	Interpretability / Key Insights	Main Results & Conclusions
(Woods et al., 2018a)	Talent Devel-opment & Competition Comparison	Classification of competition level (elite youth U20 vs. senior NRL) using 12 team performance indicators (runs, tackles, missed tackles, kicks, etc.).	CI classification tree correctly classified 79% of U20 and 93% of NRL games.	Key discriminators: 'all runs', 'tackles', 'tackle breaks', 'missed tackles', 'kicks'. NRL games = more runs and tackles, fewer missed tackles. U20 = higher tackle breaks, more errors.	Authors conclude that NRL and U20 competitions show distinct gameplay profiles. U20 players entering NRL may lack exposure to required tackling capacity and physicality. Coaches should focus on tackling ability and physical development in U20s. Suggests "bridging" via State League participation to aid transition. Practical implication: training interventions should aim to align youth gameplay with senior competition demands.
(Zhao et al., 2019)	Talent Identification & Sport-Specific Profiling	Classification of U15–U16 male athletes (basketball, fencing, judo, swimming, table tennis, volleyball) into their respective sport based on 25 tests (18 anthropometric, 5 physiological, 2 motor).	DA: 71.3% correct classification (original: 98.9%). Best: fencing 85%, volleyball 72.7%. Worst: basketball 57.1%. MLP: 71.0% correct classification (original: 99.3%). Best: volleyball 83.4%, table tennis 83.3%. Worst: basketball 20%.	Key discriminators: Anthropometry (height, shoulder width, crista width, Achilles tendon length), Motor (back strength, reaction time), Physiological (vital capacity, hemoglobin mass, resting HR). Volleyball = tall stature, strength, high lung capacity. Judo = strength, chest girth, Hb mass. Swimming = lung capacity, tendon length. Fencing = smaller chest/shoulder width. Table tennis = short lower leg length + strong back.	Authors conclude that generic test batteries of anthropometric, physiological, and motor measures can differentiate youth athletes by sport with ~70% accuracy, comparable to European studies. Findings confirm discriminative value of body size, strength, and aerobic capacity in talent ID. Basketball was hardest to classify due to small sample size. Implication: test batteries are useful for broad sport allocation, but need more sport-specific, larger-scale validation.

AE = Autoencoder; AFL = Australian Football League; APHV = Age at Peak Height Velocity; BJJ = Brazilian Jiu-Jitsu; BM = Body Mass; CNN = Convolutional Neural Network; CNN-AE-MG = Convolutional Neural Network – Autoencoder – Mixture Gaussian model; CI = Conditional Inference; DA = Discriminant Analysis; EM = Expectation–Maximization; F1 = F1-score (harmonic mean of precision and recall); FGA% = Field Goal Attempt Percentage; Hb = Hemoglobin; HR = Heart Rate; IGF-1 = Insulin-like Growth Factor 1; KNN = K-Nearest Neighbors; Lasso = Least Absolute Shrinkage and Selection Operator regression; LR = Logistic Regression; MLP = Multilayer Perceptron; nI-WAVE = Nonlinear Importance-Weighted Autoencoding Variational Inference with normalizing flow priors; NRL = National Rugby League; PCA = Principal Component Analysis; PCDEQ = Psychological Characteristics of Developing Excellence Questionnaire; PHV = Peak Height Velocity; RF = Random Forest; ROC AUC = Receiver Operating Characteristic – Area Under the Curve; SHAP = SHapley Additive exPlanations; SI = Sport Initiation; U20 = Under-20 age category; U12/U14/U18 = Under-12 / Under-14 / Under-18 age categories; YODA = Youth Online Diagnostic Assessment; YUVA-SQ = Youth Universal Value Assessment – Scouting Questionnaire.

Discussion

This systematic review synthesized evidence on the application of ML methods in sport TID and development. Across the included studies, ML was employed for diverse purposes, ranging from predicting selection and performance outcomes to supporting team formation, profiling, maturation assessment, and scouting. The findings highlight the challenges of applying ML in this domain: on one hand, advanced algorithms can capture complex, multidimensional patterns that traditional statistical approaches may overlook; on the other, the heterogeneity of data types, small sample sizes, and lack of external validation continue to limit their translational value.

This capacity to model multidimensional structure aligns closely with the ecological dynamics view of talent development, in which performance emerges from interaction-dominant rather than variable-dominant processes. ML's real strength lies not merely in detecting correlations among isolated predictors but in uncovering higher-order patterns that emerge from the interaction of biological, psychological, and environmental constraints (Reis et al., 2024). Accordingly, future research should prioritize feature sets and modeling approaches that represent these interdependent relationships - such as contextual, temporal, and relational variables - thereby aligning computational design with the ecological nature of athlete development.

Selection prediction

The synthesis of selection-focused studies demonstrates that ML models can capture important physical, technical, psychological, and socio-cultural factors associated with advancement or deselection in talent pathways. Models such as XGBoost, neural networks, and one-class SVMs achieved moderate to high predictive validity in academy soccer (Jauhiainen et al., 2019; Jennings et al., 2024; Altmann et al., 2024), while decision trees and hybrid deep learning architectures produced high accuracy in school-based settings (Theagarajan and Bhanu, 2021; Abidin and Erdem, 2025). Several studies emphasized that physical and skill-related variables (e.g., sprinting ability, countermovement jump, ball control) remain consistently influential in selection decisions, while psychological characteristics such as coping under pressure and emotional regulation also emerged as critical predictors (Owen et al., 2022; Kelly et al., 2022). Importantly, socio-cultural and relative age effects were shown to influence outcomes, underscoring that selection is not solely determined by athletic performance (Craig and Swinton, 2021; Brown et al., 2024).

Nevertheless, these studies highlight important limitations. Predictive accuracies often fell below thresholds typically required for decision-making in practice (e.g., AUC < 0.70, (Altmann et al., 2024)), while external validation was rare, raising concerns about generalizability across sports, contexts, and samples. This pattern underscores a crucial conceptual distinction between apparent validity - performance measured within the development sample - and transportable validity, which reflects how well a model generalizes to independent, real-world contexts. For example, a model predicting academy selection may achieve high internal accuracy (AUC ≈ 0.85) through resampling or cross-validation, yet when applied to a different club, season, or cohort, its performance may degrade to AUC ≈ 0.65. Such declines are not merely statistical artifacts but manifestations of the context-bound, dynamic nature of athlete development, where the distribution of constraints and opportunities shifts across settings. Recognizing this difference reinforces that external validation is not only a methodological requirement but a theoretical test of whether the modeled relationships capture genuine developmental regularities rather than local sampling patterns.

Many models also relied too much on physical test data, which limits interpretability when predicting long-term success within already selected elite groups (Craig and Swinton, 2021). Small sample sizes and imbalance between selected and deselected athletes further restrict model robustness (Jauhiainen et al., 2019). These findings emphasize that ML should not replace expert judgment but instead complement existing scouting frameworks.

Moreover, the dominance of soccer-based studies likely shapes the implicit model priors in this field, since features that are salient in invasion games (e.g., intermittent high-speed running, rapid change of direction, spatial–temporal awareness, and transition behaviors) are overrepresented in training data and outcome labels. As a result, ML models - and the feature-engineering conventions they normalize - may capture sport-specific regularities that do not readily transfer to sports with different task dynamics. This concentration can narrow ecological validity, as the performer–environment couplings and constraint sets underpinning soccer differ from those governing performance in sports such as volleyball. Expanding the evidence base beyond invasion games and encouraging cross-sport external validation would therefore strengthen the domain generalizability of ML applications in TID.

Performance prediction

Studies applying ML to performance prediction showed promising results in linking physiological and technical markers with skill-based and in-game outcomes. Early work (Cornforth et al., 2015) demonstrated that heart rate variability and environmental data could moderately predict match loads in Australian football. More recent studies (Duncan et al., 2024; Sandamal et al., 2024) expanded to youth skill assessment, where ML algorithms predicted soccer dribbling ability and test-based fitness with high accuracy when including multidimensional features such as fundamental motor skills, anthropometry, and hormonal profiles. Random Forest and XGBoost emerged as strong performers, offering predictive power and capturing nonlinear relationships in volleyball performance from anthropometric data (Sanjaykumar et al., 2024).

Despite these advances, performance prediction studies also exhibit challenges. The use of laboratory or field-test performance outcomes raises questions about ecological validity for predicting actual match performance. Furthermore, over-reliance on physiological data may neglect tactical, cognitive, and psychosocial contributors to performance. While explainable ML techniques provide interesting information into feature importance, few studies validated whether these insights align with real-world coaching expertise. To enhance translation, future work should integrate multimodal data sources and conduct prospective validation in competitive environments.

Team formation & position classification

The reviewed studies demonstrate that ML can approximate and in some cases outperform coach-derived decisions regarding position classification and team formation. For example, Random Forest and Multilayer Perceptrons achieved >90% accuracy in predicting player positions and generating lineups closely resembling coaches' choices in youth soccer (Abidin, 2021). Bayesian and tree-based models also assigned players to suitable positions with very high accuracy when using multidimensional skill ratings (Razali et al., 2017). Even when accuracy was lower, as in Australian football positional classification (Woods et al., 2018b), ML revealed meaningful patterns, such as the overlap between defenders and forwards, or the distinctiveness of midfielders.

However, most models were trained on small or academy-level datasets, limiting their generalizability across contexts. For instance in Australian football study (Woods et al., 2018b), poor classification of rucks highlighted that some roles remain underrepresented or difficult to capture with standard performance indicators. External or longitudinal validation of team formation models is virtually absent, and practical adoption will require integration with real-time data streams rather than retrospective or synthetic datasets. Thus, while ML shows strong potential in complementing coaching decisions, its utility remains contingent on larger, multi-sample validation and the inclusion of richer, role-specific features.

Profiling, development, scouting & maturation

Studies beyond direct selection and performance prediction illustrate the expanding scope of ML in talent identification and development. Morphological and neuromuscular profiling models showed value for orienting youth into appropriate sports (de Almeida-Neto et al., 2023), while cluster and outlier analyses revealed concerning early specialization patterns in basketball compared with professional norms (Contreras-García et al., 2024). Deep learning models integrating autoencoders and Gaussian mixtures provided accurate classification of youth fitness levels (Ge, 2024), while explainable ML approaches accurately predicted biological maturation status (Retzepis et al., 2024). Studies on scouting systems in women's and men's football (Venkataraman et al., 2024; López-De-Armentia, 2024) highlight the growing use of ML and automated data collection in expanding recruitment pipelines, particularly where resources are scarce. These findings underline ML's versatility in supporting orientation, development monitoring, and scouting beyond narrow predictive tasks.

Nevertheless, several limitations constrain the translation of these broader applications. Many studies remain proof-of-concept, conducted with small or single-institution datasets (de Almeida-Neto et al., 2023; Retzepis et al., 2024), or descriptive case studies without predictive validation (Venkataraman et al., 2024). External generalizability is especially limited where region-specific environmental effects or sample-specific datasets dominate (Sandamal et al., 2024; Contreras-García et al., 2024).

Limitations on ML reporting

Across the included studies, the analysis domain emerged as the most frequent source of high risk of bias, primarily due to small samples, reliance on internal validation, or use of synthetic/augmented data without adequate safeguards against optimism. For example, a study (Abidin, 2021) relied on only 21 real players supplemented with synthetic augmentation, producing very high accuracies but at the expense of validity. Similarly, another study (Abidin and Erdem, 2025) reported accuracies above 97% but did so without external validation and with imbalanced data, leaving open the possibility of overfitting. Even in larger, better-resourced settings (Altmann et al., 2024), while participants and predictors were appropriately defined, the lack of calibration and external testing led to an overall "unclear" rating in the analysis domain. These aspects suggest that although predictive modeling is advancing in youth TID research, methodological rigor in handling imbalance, avoiding leakage, and validating models externally is still uncommon.

A second recurrent issue relates to applicability of predictors and outcomes, especially where subjective or indirect measures were used. For instance, studies using coach-rated assessments as input variables (Abidin, 2021; Abidin and Erdem, 2025) faced concerns that these subjective scores could embed bias or even overlap with the outcome being predicted. Other study (de Almeida-Neto et al., 2023) used cross-sport orientation outcomes rather than within-sport selection, which limited the direct applicability of their findings to talent identification in team sports. In contrast, where predictors were standardized and outcomes were objectively defined (Craig and Swinton, 2021), the risk of bias was lower, even if model performance was weak. Overall, most included studies were judged at least "some concern" for applicability, underscoring that future work should prioritize transparent, objective measures aligned closely with actual selection or progression outcomes.

Limitations of this systematic review, future research and practical applications

This review has limitations that should be acknowledged. Despite a comprehensive search and systematic screening process, it is possible that relevant studies were missed, particularly those published in grey literature (e.g., technical reports, theses). The exclusion of grey literature was a deliberate methodological choice to maintain peer-reviewed quality standards; however, it introduces the possibility of publication bias, as studies reporting weaker or non-significant results are less likely to appear in indexed journals. Consequently, the synthesized evidence may overrepresent positive findings and potentially overestimate ML model performance. This limitation may be important, as it reflects a broader tendency within data-driven research toward selective visibility of success - a phenomenon that underscores the need for greater transparency, data sharing, and preregistration in ML-based sports science. Moreover, the heterogeneity of sports, outcome measures, and machine learning approaches precluded meta-analysis and restricted the synthesis to a structured narrative. The reliance on published results also meant that incomplete reporting of performance metrics or validation methods could not be clarified or supplemented, further limiting interpretability. Finally, as many included studies were exploratory, single-sample, or lacked external validation, the evidence base summarized here represents an emerging rather than mature field.

Interpretability emerged as one of the least consistently addressed dimensions across studies, yet it represents a continuum of conceptual transparency rather than a binary property. At the most basic level, interpretability can involve global feature importance or coefficient-based rankings that indicate which variables most influence predictions. More advanced methods, such as SHAP (SHapley Additive Explanations) or LIME (Local Interpretable Model-Agnostic Explanations), allow for instance-level attribution, showing how specific inputs contribute to individual outcomes. At the highest tier, counterfactual reasoning provides actionable insight by simulating how changes in certain features might alter selection probabilities or developmental trajectories. Viewing interpretability hierarchically underscores that transparency in ML is scalable—from descriptive feature inspection to causal exploration—

and that its depth should align with the practical stakes of decision-making in TID.

Looking ahead, future research should prioritize larger, longitudinal, and multi-sport datasets that allow for robust model development and both statistical and ecological external validation. In addition to conventional hold-out or cross-cohort testing, ecological external validation involves evaluating model performance across different clubs, regions, and competition levels to ensure contextual robustness and ecological realism. Such cross-setting validation helps determine whether predictive patterns reflect genuine developmental principles or context-specific artifacts, bridging methodological rigor with the complex, adaptive nature of sport environments. Standardized reporting of ML pipelines - including feature engineering, calibration assessment, validation strategies, and interpretability methods - would improve transparency and comparability across studies. Greater integration of multidimensional data is also needed to capture the complexity of talent development. Moreover, collaboration between sport scientists, data scientists, and practitioners will be essential to ensure that models are not only accurate but also interpretable, ethically sound, and practically relevant. By embracing open science practices and methodological rigor, the field can move beyond optimism bias toward a more cumulative, self-correcting body of evidence that meaningfully informs talent identification and development systems.

To enhance reproducibility and comparability, future ML studies in talent identification should adopt, at minimum, clearly describe their data partitioning strategy, including whether splits were performed at the athlete or trial level; outline steps for leakage control to prevent information overlap between training and testing sets; report how class imbalance was handled within validation folds; and include both discrimination and calibration metrics (e.g., AUC, Brier score, calibration slope). In addition, transparency around fairness auditing - such as assessing model performance across relative-age quartiles, sex, or maturation status - will improve interpretability and ethical accountability. Consistent reporting of these elements would substantially strengthen the methodological quality, transparency, and applied trustworthiness of ML research in youth talent identification.

To promote equitable predictions across subpopulations, we propose a minimal fairness framework specifying main covariates that should be recorded, modeled, and audited in youth TID, as exemples, birth quarter/relative age, biological maturation status (e.g., PHV indicators), and socio-economic background (e.g., school type or deprivation index), alongside sex and playing context (e.g., region/club resource level). These variables should be (i) pre-specified in protocols, (ii) considered as features or stratification factors where appropriate, and (iii) subjected to subgroup and intersectional audits reporting discrimination, calibration, and error-rate parity at a stated operating point. If disparities are detected, studies should apply bias-mitigation procedures (e.g., reweighting, stratified sampling, threshold adjustment, post-hoc recalibration) and re-report subgroup metrics.

From a practical standpoint, the findings of this review suggest that ML may have potential to complement, rather than replace, traditional talent identification and development practices. Current evidence indicates that ML models can highlight patterns across large, multidimensional datasets and may assist coaches and scouts in refining their decisions or monitoring athlete development. However, given the frequent limitations of small sample sizes, context-specific data, and limited external validation, these tools should be viewed as exploratory decision-support aids rather than definitive selection instruments. Practitioners are advised to use ML outputs in conjunction with expert judgment, holistic evaluation of athletes, and awareness of potential biases (e.g., relative age, socio-cultural influences). This complementary role can be understood along two interconnected pathways, namely an operational pathway, in which ML assists practitioners with data-driven screening, workload monitoring, and early flagging of developmental trends to enhance decision efficiency, and a discovery pathway, where ML identifies novel, interaction-based patterns among physical, technical, and psychosocial constraints that can inform longitudinal experimentation and theory development. These pathways illustrate that the value of ML lies not in replacing human expertise but in augmenting it - bridging empirical discovery with applied decision-making in youth talent systems. Careful integration in practice may enhance efficiency and provide additional perspectives, but overreliance on unvalidated models risks reinforcing existing inequalities or producing misleading conclusions.

To operationalize these findings, practitioners could adopt tiered decision protocols in which ML models are first used for broad early screening - prioritizing high sensitivity to avoid missing potential talent - followed by structured expert evaluation emphasizing context, adaptability, and psychosocial maturity. Such hybrid frameworks can combine algorithmic efficiency with human interpretive depth, ensuring that automated outputs inform but do not dictate selection. In this way, ML functions as an evidence-based triage tool that supports individualized monitoring, facilitates ongoing re-evaluation, and helps direct coaching resources toward athletes with emerging potential rather than early advantage.

From a practitioner perspective, the implementation of ML in TID can also be conceptualized as a sequential decision pathway encompassing model development, validation, deployment, and monitoring. During development, multidisciplinary teams should ensure data representativeness, apply rigorous leakage control, and use nested cross-validation to optimize model tuning. Validation should progress from internal to independent external testing to evaluate transportability and calibration before any operational use. In deployment, ML outputs should serve as decision-support tools within structured selection frameworks - for instance, as high-sensitivity screening aids that prompt subsequent expert evaluation. Finally, ongoing monitoring is essential to detect model drift, reassess fairness across athlete subgroups, and recalibrate performance metrics as data and populations evolve. This cyclical process ensures that ML models remain methodologically sound, contextually relevant, and ethically aligned with the developmental principles of youth sport.

Conclusion

This systematic review found that research applying ML in sport talent identification remains limited in scope but expanding. The majority of available studies focused on selection prediction tasks, particularly in soccer and other team sports, where algorithms were used to forecast admission, progression, or draft success. A smaller but growing body of work addressed performance prediction, leveraging physiological, anthropometric, or cognitive markers to estimate test results or in-game performance. Fewer studies explored team formation and positional classification, and an emerging set of contributions examined broader applications such as profiling, maturation, and scouting support. Across domains, Random Forest, gradient boosting methods, and neural networks were the most frequently applied, often achieving moderate to high internal accuracy. However, very few studies provided external validation, and most were conducted on relatively small, single-sport or academy-specific datasets, limiting generalizability.

The findings suggest that while ML offers clear potential to enrich talent identification and development systems, its current role should be viewed as exploratory and complementary rather than decisive. The predominance of selection-focused studies highlights a narrow evidence base, with underrepresentation of longitudinal designs, female athletes, and diverse sporting contexts. Moreover, interpretability methods - although increasingly adopted - remain inconsistently applied, and socio-cultural or psychological factors are still less frequently integrated than physical and technical measures. Future progress will depend on larger, multi-sample datasets, standardized reporting of algorithms and metrics, and collaborative efforts to embed interpretability and equity within predictive pipelines. Until such methodological and theoretical maturity is achieved, the use of ML in practice should remain cautious, serving as a support to - not a substitute for - expert judgment and holistic athlete evaluation. Ultimately, in youth TID, transparency, transportability, and theoretical coherence are the pillars upon which meaningful ML applications must be built.

Acknowledgements

This study was supported by the Project of China West Normal University, Project Number: [CWNUJG2024098]. The author reports no actual or potential conflicts of interest. The datasets generated and analyzed in this study are not publicly available, but are available from the corresponding author who organized the study upon reasonable request. All experimental procedures were conducted in compliance with the relevant legal and ethical standards of the country where the study was performed.

References

Abidin, D. (2021). A case study on player selection and team formation in football with machine learning. *Turkish Journal of Electrical Engineering and Computer Sciences* **29**, 1672-1691. https://doi.org/10.3906/elk-2005-27

Abidin, D. and Erdem, M. G. (2025). SCM-DL: Split-Combine-Merge Deep Learning Model Integrated With Feature Selection in Sports for Talent Identification. *IEEE Access* **13**, 71148-71172. https://doi.org/10.1109/ACCESS.2025.3562551

Altmann, S., Ruf, L., Thiem, S., Beckmann, T., Wohak, O., Romeike, C. and Härtel, S. (2024). Prediction of talent selection in elite male youth soccer across 7 seasons: A machine-learning approach. *Journal of Sports Sciences*, 1-14. https://doi.org/10.1080/02640414.2024.2442850

Barraclough, S., Till, K., Kerr, A. and Emmonds, S. (2022). Methodological approaches to talent identification in team sports: A narrative review. *Sports* **10**, 81. https://doi.org/10.3390/sports10060081

Beato, M., Jaward, M. H., Nassis, G. P., Figueiredo, P., Clemente, F. M. and Krustrup, P. (2025). An educational review on machine learning: A SWOT analysis for implementing machine learning techniques in football. *International Journal of Sports Physiology and Performance* **20**, 183-191. https://doi.org/10.1123/ijspp.2024-0247

Brown, T., Cook, R., Gough, L. A., Khawaja, I., McAuley, A. B. T. and Kelly, A. L. (2024). Exploring the multidimensional characteristics of selected and non-selected White British and British South Asian youth cricketers: An exploratory machine learning approach. *Youth* **4**, 718-734. https://doi.org/10.3390/youth4020048

Chmait, N. and Westerbeek, H. (2021). Artificial intelligence and machine learning in sport research: An introduction for non-data scientists. *Frontiers in Sports and Active Living* **3**. https://doi.org/10.3389/fspor.2021.682287

Cobley, S., Baker, J., Wattie, N. and McKenna, J. (2009). Annual age-grouping and athlete development: A meta-analytical review of relative age effects in sport. *Sports Medicine* **39**, 235-256. https://doi.org/10.2165/00007256-200939030-00005

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., Boulesteix, A.-L., Camaradou, J. C., Celi, L. A., Denaxas, S., Denniston, A. K., Glocker, B., Golub, R. M., Harvey, H., Heinze, G., Hoffman, M. M., Kengne, A. P., Lam, E., Lee, N., Loder, E. W., Maier-Hein, L., Mateen, B. A., McCradden, M. D., Oakden-Rayner, L., Ordish, J., Parnell, R., Rose, S., Singh, K., Wynants, L. and Logullo, P. (2024). TRIPOD+AI statement: Updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ*, e078378. https://doi.org/10.1136/bmj-2023-078378

Contreras-García, J. M., Courel-Ibáñez, J., Piñar-López, M. I. and Ibáñez, S. J. (2024). Early specialization in formative basketball: A machine learning analysis of shooting patterns in U14 and professional players. *Journal of Sports Sciences*, 1-8. https://doi.org/10.1080/02640414.2024.2436809

Cordeiro, M. C., Cathain, C. O., Daly, L., Kelly, D. T. and Rodrigues, T. B. (2025). A synthetic data-driven machine learning approach for athlete performance attenuation prediction. *Frontiers in Sports and Active Living* **7**. https://doi.org/10.3389/fspor.2025.1607600

Cornforth, D., Campbell, P., Nesbitt, K., Robinson, D. and Jelinek, H. F. (2015). Prediction of game performance in Australian football using heart rate variability measures. *International Journal of Signal and Imaging Systems Engineering* **8**, 80-88. https://doi.org/10.1504/IJSISE.2015.067072

Craig, T. P. and Swinton, P. (2021). Anthropometric and physical performance profiling does not predict professional contracts awarded in an elite Scottish soccer academy over a 10-year period. *European Journal of Sport Science* **21**, 1101-1110. https://doi.org/10.1080/17461391.2020.1808079

de Almeida-Neto, P. F., Neto, R. B., de Matos, D. G., de Medeiros, J. A., Bulhões-Correia, A., Jeffreys, I., Lobato, C. H., Aidar, F. J., Dantas, P. M. S. and Cabral, B. G. A. T. (2023). Using artificial neural networks to help in the process of sports selection and orientation through morphological and biodynamic parameters: A pilot study. *Sport Sciences for Health* **19**, 929-937. https://doi.org/10.1007/s11332-022-00986-1

Deprez, D., Buchheit, M., Fransen, J., Pion, J., Lenoir, M., Philippaerts, R. M. and Vaeyens, R. (2015). A longitudinal study investigating the stability of anthropometry and soccer-specific endurance in pubertal high-level youth soccer players. *Journal of Sports Science and Medicine* **14**, 418-426. https://pubmed.ncbi.nlm.nih.gov/25983593/

Duncan, M. J., Eyre, E. L. J., Clarke, N., Hamid, A. and Jing, Y. (2024). Importance of fundamental movement skills to predict technical skills in youth grassroots soccer: A machine learning approach.

International Journal of Sports Science & Coaching **19**, 1042-1049. https://doi.org/10.1177/17479541231202015
Finnegan, L., van Rijbroek, M., Oliva-Lozano, J. M., Cost, R. and Andrew, M. (2024). Relative age effect across the talent identification process of youth female soccer players in the United States: Influence of birth year, position, biological maturation, and skill level. *Biology of Sport* **41**, 241-251. https://doi.org/10.5114/biolsport.2024.136085
Formenti, D., Trecroci, A., Duca, M., Vanoni, M., Ciovati, M., Rossi, A. and Alberti, G. (2022). Volleyball-specific skills and cognitive functions can discriminate players of different competitive levels. *Journal of Strength and Conditioning Research* **36**, 813-819. https://doi.org/10.1519/JSC.0000000000003519
Gallitto, G., Englert, R., Kincses, B., Kotikalapudi, R., Li, J., Hoffschlag, K., Bingel, U. and Spisak, T. (2025). External validation of machine learning models—registered models and adaptive sample splitting. *GigaScience* **14**. https://doi.org/10.1093/gigascience/giaf036
Ge, C. (2024). Optimization study of a dynamic assessment model of physical fitness for youth basketball training. *Applied Mathematics and Nonlinear Sciences* **9**. https://doi.org/10.2478/amns-2024-3396
Gogos, B. J., Larkin, P., Haycraft, J. A. Z., Collier, N. F. and Robertson, S. (2020). Combine performance, draft position and playing position are poor predictors of player career outcomes in the Australian Football League. *PLOS ONE* **15**, e0234400. https://doi.org/10.1371/journal.pone.0234400
Güllich, A. (2014). Selection, de-selection and progression in German football talent promotion. *European Journal of Sport Science* **14**, 530-537. https://doi.org/10.1080/17461391.2013.858371
Haan, M. de, van der Zwaard, S., Sanders, J., Beek, P. J. and Jaspers, R. T. (2025). Beyond playing positions: Categorizing soccer players based on match-specific running performance using machine learning. *Journal of Sports Science and Medicine*, 565-577. https://doi.org/10.52082/jssm.2025.565
Herrebrøden, H. and Bjørndal, C. T. (2022). Youth international experience is a limited predictor of senior success in football: The relationship between U17, U19, and U21 experience and senior elite participation across nations and playing positions. *Frontiers in Sports and Active Living* **4**. https://doi.org/10.3389/fspor.2022.875530
Jamil, M., Phatak, A., Mehta, S., Beato, M., Memmert, D. and Connor, M. (2021). Using multiple machine learning algorithms to classify elite and sub-elite goalkeepers in professional men's football. *Scientific Reports* **11**, 22703. https://doi.org/10.1038/s41598-021-01187-5
Jauhiainen, S., Äyrämö, S., Forsman, H. and Kauppi, J.-P. (2019). Talent identification in soccer using a one-class support vector machine. *International Journal of Computer Science in Sport* **18**, 125-136. https://doi.org/10.2478/ijcss-2019-0021
Jennings, J., Perrett, J. C., Wundersitz, D. W., Sullivan, C. J., Cousins, S. D. and Kingsley, M. I. (2024). Predicting successful draft outcome in Australian rules football: Model sensitivity is superior in neural networks when compared to logistic regression. *Plos One* **19**, e0298743. https://doi.org/10.1371/journal.pone.0298743
de Jong, Y., Ramspek, C. L., Zoccali, C., Jager, K. J., Dekker, F. W. and van Diepen, M. (2021). Appraising prediction research: A guide and meta-review on bias and applicability assessment using the Prediction model Risk Of Bias ASsessment Tool (PROBAST). *Nephrology* **26**, 939-947. https://doi.org/10.1111/nep.13913
Kapoor, S. and Narayanan, A. (2023). Leakage and the reproducibility crisis in machine-learning-based science. *Patterns* **4**, 100804. https://doi.org/10.1016/j.patter.2023.100804
Kelly, A. L., Williams, C. A., Cook, R., Sáiz, S. L. J. and Wilson, M. R. (2022). A multidisciplinary investigation into the talent development processes at an English football academy: A machine learning approach. *Sports* **10**, 159. https://doi.org/10.3390/sports10100159
Kilian, P., Leyhr, D., Urban, C. J., Höner, O. and Kelava, A. (2023). A deep learning factor analysis model based on importance-weighted variational inference and normalizing flow priors: Evaluation within a set of multidimensional performance assessments in youth elite soccer players. *Statistical Analysis and Data Mining: The ASA Data Science Journal* **16**, 474-487. https://doi.org/10.1002/sam.11632
Larkin, P. and O'Connor, D. (2017). Talent identification and recruitment in youth soccer: Recruiter's perceptions of the key attributes for player recruitment. *PLOS ONE* **12**, e0175716. https://doi.org/10.1371/journal.pone.0175716
Leckey, C., van Dyk, N., Doherty, C., Lawlor, A. and Delahunt, E. (2025). Machine learning approaches to injury risk prediction in sport: A scoping review with evidence synthesis. *British Journal of Sports Medicine* **59**, 491-500. https://doi.org/10.1136/bjsports-2024-108576
López-De-Armentia, J. (2024). WTDTool: Women's talent detection tool. In *2024 IEEE International Workshop on Sport, Technology and Research (STAR)* (pp. 144-149). IEEE. https://doi.org/10.1109/STAR62027.2024.10635969
Malina, R. M., Rogol, A. D., Cumming, S. P., Coelho E Silva, M. J. and Figueiredo, A. J. (2015). Biological maturation of youth athletes: Assessment and implications. *British Journal of Sports Medicine* **49**, 852-859. https://doi.org/10.1136/bjsports-2015-094623
Nassis, G., Verhagen, E., Brito, J., Figueiredo, P. and Krustrup, P. (2023). A review of machine learning applications in soccer with an emphasis on injury risk. *Biology of Sport* **40**, 233-239. https://doi.org/10.5114/biolsport.2023.114283
Owen, J., Owen, R., Hughes, J., Leach, J., Anderson, D. and Jones, E. (2022). Psychosocial and physiological factors affecting selection to regional age-grade rugby union squads: A machine learning approach. *Sports* **10**, 35. https://doi.org/10.3390/sports10030035
Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., McGuinness, L. A., Stewart, L. A., Thomas, J., Tricco, A. C., Welch, V. A., Whiting, P. and Moher, D. (2021a). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, n71. https://doi.org/10.1136/bmj.n71
Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., McGuinness, L. A., Stewart, L. A., Thomas, J., Tricco, A. C., Welch, V. A., Whiting, P. and Moher, D. (2021b). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *Systematic Reviews* **10**, 89. https://doi.org/10.1186/s13643-021-01626-4
Ravé, G., Granacher, U., Boullosa, D., & Hackney, A. C. (2020). How to use global positioning systems (GPS) data to monitor training load in the "real world" of elite soccer. *Frontiers in Physiology*, **11**, 944. https://doi.org/10.3389/fphys.2020.00944
Razali, N., Mustapha, A., Yatim, F. A. and Ab Aziz, R. (2017). Predicting player position for talent identification in association football. *IOP Conference Series: Materials Science and Engineering* **226**, 012087. https://doi.org/10.1088/1757-899X/226/1/012087
Reilly, T., Williams, A. M., Nevill, A. and Franks, A. (2000). A multidisciplinary approach to talent identification in soccer. *Journal of Sports Sciences* **18**, 695-702. https://doi.org/10.1080/02640410050120078
Reis, F. J. J., Alaiti, R. K., Vallio, C. S. and Hespanhol, L. (2024). Artificial intelligence and machine learning approaches in sports: Concepts, applications, challenges, and future perspectives. *Brazilian Journal of Physical Therapy* **28**, 101083. https://doi.org/10.1016/j.bjpt.2024.101083
Retzepis, N.-O., Avloniti, A., Kokkotis, C., Protopapa, M., Stampoulis, T., Gkachtsou, A., Pantazis, D., Balampanos, D., Smilios, I. and Chatzinikolaou, A. (2024). Identifying key factors for predicting the age at peak height velocity in preadolescent team sports athletes using explainable machine learning. *Sports* **12**, 287. https://doi.org/10.3390/sports12110287
Richardson, E., Trevizani, R., Greenbaum, J. A., Carter, H., Nielsen, M. and Peters, B. (2024). The receiver operating characteristic curve accurately assesses imbalanced datasets. *Patterns* **5**, 100994. https://doi.org/10.1016/j.patter.2024.100994
Rico-González, M., Pino-Ortega, J., Méndez, A., Clemente, F. and Baca, A. (2023). Machine learning application in soccer: A systematic review. *Biology of Sport* **40**, 249-263. https://doi.org/10.5114/biolsport.2023.112970
Sandamal, K., Arachchi, S., Erkudov, V. O., Rozumbetov, K. U. and Rathnayake, U. (2024). Explainable artificial intelligence for

fitness prediction of young athletes living in unfavorable environmental conditions. *Results in Engineering* **23**, 102592. https://doi.org/10.1016/j.rineng.2024.102592
Sanjaykumar, S., Natarajan, S., Lakshmi, P. Y., Kalmykova, Y., Lobo, J., Pavlović, R. and Setiawan, E. (2024). Machine learning analysis for predicting performance in female volleyball players in India. *Journal of Human Sport and Exercise* **20**, 207-215. https://doi.org/10.55860/cn2vdj44
Sarmento, H., Anguera, M. T., Pereira, A. and Araújo, D. (2018). Talent identification and development in male football: A systematic review. *Sports Medicine* **48**, 907-931. https://doi.org/10.1007/s40279-017-0851-7
Seifert, L., Araújo, D., Komar, J. and Davids, K. (2017). Understanding constraints on sport performance from the complexity sciences paradigm: An ecological dynamics framework. *Human Movement Science* **56**, 178-180. https://doi.org/10.1016/j.humov.2017.05.001
Seifert, L., Hacques, G. and Komar, J. (2022). The ecological dynamics framework: An innovative approach to performance in extreme environments: A narrative review. *International Journal of Environmental Research and Public Health* **19**, 2753. https://doi.org/10.3390/ijerph19052753
Smith, K. L., Weir, P. L., Till, K., Romann, M. and Cobley, S. (2018). Relative age effects across and within female sport contexts: A systematic review and meta-analysis. *Sports Medicine* **48**, 1451-1478. https://doi.org/10.1007/s40279-018-0915-3
Theagarajan, R. and Bhanu, B. (2021). An automated system for generating tactical performance statistics for individual soccer players from videos. *IEEE Transactions on Circuits and Systems for Video Technology* **31**, 632-646. https://doi.org/10.1109/TCSVT.2020.2982580
Till, K. and Baker, J. (2020). Challenges and [possible] solutions to optimizing talent identification and development in sport. *Frontiers in Psychology* **11**. https://doi.org/10.3389/fpsyg.2020.00664
Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine* **25**, 44-56. https://doi.org/10.1038/s41591-018-0300-7
Vabalas, A., Gowen, E., Poliakoff, E. and Casson, A. J. (2019). Machine learning algorithm validation with a limited sample size. *PLOS ONE* **14**, e0224365. https://doi.org/10.1371/journal.pone.0224365
Vaeyens, R., Lenoir, M., Williams, A. M. and Philippaerts, R. M. (2008). Talent identification and development programmes in sport. *Sports Medicine* **38**, 703-714. https://doi.org/10.2165/00007256-200838090-00001
Varghese, M., Ruparell, S. and LaBella, C. (2022). Youth athlete development models: A narrative review. *Sports Health: A Multidisciplinary Approach* **14**, 20-29. https://doi.org/10.1177/19417381211055396
Venkataraman, S., Sundharakumar, K., Bharathi Malakreddy, A. and Natarajan, S. (2024). YUVA-SQ: A cognitive scouting model for the beautiful game. In *2024 5th International Conference on Innovative Trends in Information Technology (ICITIIT)* (pp. 1-6). IEEE. https://doi.org/10.1109/ICITIIT61487.2024.10580784
Williams, A. M. and Reilly, T. (2000). Talent identification and development in soccer. *Journal of Sports Sciences* **18**, 657-667. https://doi.org/10.1080/02640410050120041
Wolff, R. F., Moons, K. G. M., Riley, R. D., Whiting, P. F., Westwood, M., Collins, G. S., Reitsma, J. B., Kleijnen, J. and Mallett, S. (2019). PROBAST: A tool to assess the risk of bias and applicability of prediction model studies. *Annals of Internal Medicine* **170**, 51-58. https://doi.org/10.7326/M18-1376
Woods, C. T., McKeown, I., Rothwell, M., Araújo, D., Robertson, S. and Davids, K. (2020). Sport practitioners as sport ecology designers: How ecological dynamics has progressively changed perceptions of skill "acquisition" in the sporting habitat. *Frontiers in Psychology* **11**. https://doi.org/10.3389/fpsyg.2020.00654
Woods, C. T., Robertson, S., Sinclair, W. H., Till, K., Pearce, L. and Leicht, A. S. (2018a). A comparison of game-play characteristics between elite youth and senior Australian National Rugby League competitions. *Journal of Science and Medicine in Sport* **21**, 626-630. https://doi.org/10.1016/j.jsams.2017.10.003
Woods, C. T., Veale, J., Fransen, J., Robertson, S. and Collier, N. F. (2018b). Classification of playing position in elite junior Australian football using technical skill indicators. *Journal of Sports Sciences* **36**, 97-103. https://doi.org/10.1080/02640414.2017.1282621
Zhao, K., Hohmann, A., Chang, Y., Zhang, B., Pion, J. and Gao, B. (2019). Physiological, anthropometric, and motor characteristics of elite Chinese youth athletes from six different sports. *Frontiers in Physiology* **10**. https://doi.org/10.3389/fphys.2019.00405

Key points

- Machine learning (ML) can identify complex talent patterns across physical, technical, and psychological data, but it should complement—not replace—expert judgment.
- Most studies show moderate accuracy but lack external validation, making their generalizability and real-world reliability limited.
- Current research is constrained by small samples and bias, highlighting the need for larger, multi-sport, and longitudinal datasets with standardized reporting and validation.

AUTHOR BIOGRAPHY

Qingrong TANG
Employment
Geely University of China, Chengdu, 641400, China
Degree
PhD
Research interests
Artificial Intelligence., etc.
E-mail: tangqr19@gmail.com

Xiufang WEI
Employment
College of physical education and health, China West Normal University, Nan-chong, 637002, China
Degree
M.Ed.
Research interests
Physical education, etc.
E-mail: 15328439955@163.com

Bo TAN
Employment
Geely University of China, Chengdu, 641400, China
Degree
PhD
Research interests
Sports training, sports psychology, etc.
E-mail: 920514879@qq.com

✉ **Xiufang Wei**
College of physical education and health, China West Normal University, Nan-chong, 637002, China