International Journal of Yoga, Physiotherapy and Physical Education
www.sportsjournal.in
ISSN: 2456-5067
Received: 25-07-2026, Accepted: 24-08-2026, Published: 09-09-2026
Volume 11, Issue 3, 2026, Page No. 89-97

The role of artificial intelligence in promoting fairness and transparency in athlete selection: A thematic review

Goldy[1], Rajeev Kumar[1], Dr. Alok Kumar Singh[2]
Research Scholar, Department of Physical Education, Indira Gandhi National Tribal University, Amarkantak, Madhya Pradesh, India
Assistant Professor, Department of Physical Education, Indira Gandhi National Tribal University, Amarkantak, Madhya Pradesh, India
DOI: https://doi.org/10.66856/ijyppe.2026.11.3.11067

Abstract
Athlete selection and talent identification (TID) in elite sports are undergoing a paradigm shift from traditional subjective human evaluations to objective, data-driven systems powered by artificial intelligence (AI) and machine learning (ML). This thematic review paper provides a comprehensive, critical synthesis of the current literature (2020–2026) surrounding the integration of AI in sports selection systems. By analyzing key peer-reviewed studies and official project guidelines, this paper systematically explores: (1) the limitations of conventional scouting including relative age effects (RAE), maturity-selection bias, and cognitive heuristics; (2) state-of-the-art computational approaches, such as the Split-Combine-Merge Deep Learning (SCM-DL) model, LightGBM classification with SHAP interpretability, and Bayesian pattern recognition; (3) the ethical tensions between algorithmic fairness, data privacy (GDPR/HIPAA compliance), and athlete rights; and (4) human-AI collaboration. Furthermore, this review addresses the applied sports context of India, specifically examining the technological framework of the Khelo India Rising Talent Identification (KIRTI) initiative, alongside empirical machine learning research in Indian volleyball and cricket. We propose a holistic conceptual framework of an ethically aligned, transparent AI-assisted athlete selection pipeline that integrates data analytics, explainability, fairness audits, and qualitative human expertise. This review underscores that while AI can significantly enhance standardization, consistency, and analytical efficiency, it is not an automated guarantor of fairness. Responsible deployment requires continuous auditing of training datasets, explainable models, and strict human oversight to prevent the replication of historical biases and safeguard athlete welfare.

Keywords: Artificial intelligence, athlete selection, talent identification, algorithmic fairness, explainable AI (XAI), KIRTI, Sports Science

Introduction
The landscape of high-performance and Olympic sport is more competitive than ever, driven by substantial financial, commercial, and political incentives, which push national organizations to optimize athletic pathways (Chmait & Westerbeek, 2021) [4]. To manage limited resources and maximize success, national governing bodies and professional sports clubs invest heavily in Talent Identification and Development Systems (TIDS) (Till & Baker, 2020) [28]. However, traditional athlete selection processes have historically relied on subjective assessments, unweighted empirical evaluations, and informal "gut feel" heuristics of coaches and scouts (Johnston *et al*., 2025) [9]. These conventional practices are heavily prone to high error rates, cognitive biases, and systematic inequalities, such as the relative age effect (RAE) and biological maturation selection bias, which compromise selection objectivity and long-term pathway fairness (Barraclough *et al*., 2022; Kelly *et al*., 2022) [2, 10]. This subjective model leads to high rates of de-selection disputes and athlete dropouts.

In response to these limitations, the rapid "datafication" of sport accelerated by wearable technologies, inertial measurement units (IMUs), GPS-tracking, and high-speed computer vision systems has generated high-volume, multidimensional performance datasets (Li *et al*., 2025; Munoz-Macho *et al*., 2024) [1, 17]. This surge of performance data has motivated the adoption of artificial intelligence (AI) and machine learning (ML) techniques to transition sports recruitment into an objective, standardized, and evidence-based domain (Chmait & Westerbeek, 2021; Tang *et al*., 2026) [4, 27]. Artificial intelligence has also been successfully applied to enhance scoring accuracy and fairness in sports competitions, removing human subjectivity in technical evaluations (Fu & Wu, 2024) [16]. State-of-the-art ML algorithms, including gradient boosting machines (e.g., LightGBM, XGBoost) and deep neural networks, excel at processing high-dimensional, non-linear relationships across physiological, biomechanical, and cognitive variables, providing decision support that outperforms linear statistical models (Abidin & Erdem, 2025; Mănescu & Mănescu, 2025) [1].

However, the rapid implementation of AI in sports selection does not automatically guarantee fairness (Kim *et al*., 2025) [11]. Machine learning models are intrinsically retrospective, trained on historical selection data that often reflects systemic societal, racial, and gender-based discrimination (Pagano *et al*., 2022; Pessach & Shmueli, 2020) [19, 20]. Opaque, "black-box" architectures present a critical transparency challenge; if coaches and athletes cannot interpret the rationale behind a selection decision, algorithmic trust collapses, and administrative or legal disputes inevitably arise (Kranzinger *et al*., 2026) [12]. Furthermore, the continuous biometric surveillance of young athletes raises severe ethical concerns regarding data privacy, informed consent, and the de-selection risks of predictive modeling (Johnston *et al*., 2025; Kim *et al*., 2025) [15].

Research Gaps and Major Contributions of This Review

Despite the growing body of literature on sports analytics, substantial research gaps remain at the intersection of athletic selection and computer science. Existing reviews typically focus on narrow technical model accuracy in specific, high-exposure sports like professional soccer, or discuss general ethical AI principles without integrating domain-specific physical education constraints. This thematic review paper addresses these critical omissions by offering three novel contributions: (1) we systematically bridge the gap between high-dimensional machine learning architectures and applied sports science taxonomy; (2) we critically examine the mathematical codification of relative age effects and biological maturation within AI training loops, moving beyond descriptive bias accounts; and (3) we propose a holistic, human-in-the-loop conceptual framework that maps explainable AI (XAI) and algorithmic fairness audits directly onto a cyclical athlete development pathway.

Athlete Selection and Talent Identification

To establish a rigorous foundation, it is critical to clarify the taxonomy and nomenclature of the sporting pathway. As highlighted by Zhao *et al*. (2024) and Till and Baker (2020) [28, 30], talent development systems consist of four distinct, interconnected phases: (1) *Talent Detection (TAD)*, which refers to the discovery of potential performers who are not currently participating in the sport in question; (2) *Talent Identification (TID)*, which is the process of recognizing current sports participants who possess the underlying biological, psychological, and social characteristics to progress to elite-level senior status (Johnston *et al*., 2025) [9]; (3) *Talent Development (TDE)*, which provides an optimal nurturing and training environment to assist athletes in realizing their potential; and (4) *Talent Selection (TSE)*, which is the ongoing, contextualized process of identifying players at various developmental stages who demonstrate prerequisite performance levels to be selected for a specific squad, competition, or representative task.

Zhao *et al*. (2024) [30] emphasize that TID and TSE are frequently conflated, yet their operational definitions differ fundamentally. TID is inherently prognostic and long-term, aiming to predict adult potential from adolescent profiles under highly unstable, non-linear developmental constraints (Johnston *et al*., 2025) [9]. Conversely, TSE focuses on the short-term selection of an athlete's current capacity to complete a specific task, such as winning a match or completing a representative squad (Zhao *et al*., 2024) [30]. Traditional TIDS are severely limited because they attempt to predict long-term potential (TID) based on transient, short-term youth performance (TSE) (Till & Baker, 2020) [28]. This "resulting" bias evaluates the quality of a selection decision solely on the immediate outcome rather than the rigor of the decision-making process (Johnston *et al*., 2025) [9].

Because athletic development is non-optimal, non-linear, and emergent, single-timepoint cross-sectional "snapshots" fail to capture developmental trajectories, leading to high dropout and talent wastage (Barraclough *et al*., 2022) [2]. To bridge this gap, modern sports science advocates for longitudinal, multi-disciplinary profiling that integrates physical, technical, and psychosocial attributes, as emphasized by Johnston *et al*. (2025) [9]. It is within this complex, high-dimensional web of variables that AI-driven predictive modeling has emerged as a disruptive technological intervention, allowing practitioners to analyze non-linear progression curves and predict long-term trajectories with greater precision (Abidin & Erdem, 2025; Mănescu & Mănescu, 2025) [1].

Review Method

This paper employs a systematic thematic review methodology to synthesize and critically appraise current evidence on AI applications in athlete selection, in accordance with the PRISMA recommendations (Tang *et al*., 2026) [27]. A systematic search of digital databases (including Web of Science, Scopus, and PubMed) was evaluated using the population, concept, context (PCC) framework to define study eligibility, as recommended in systematic scoping guidelines (Kranzinger *et al*., 2026). Rico-González *et al*. (2023) [12, 24] conducted a systematic review highlighting the rapid growth of machine learning applications in soccer, focusing on technical and physical performance metrics. Eligible studies included peer-reviewed quantitative research and scoping reviews exploring ML and AI methods applied to athlete talent identification, predictive modeling, or sports ethics (Kim *et al*., 2025; Tang *et al*., 2026) [1, 27].

The primary search strategy combined terms such as "machine learning," "artificial intelligence," "talent identification," and "player selection" across both individual and team sports (Tang *et al*., 2026) [27]. After duplicate removal and rigorous screening, a final corpus of 25 highly relevant studies and official project materials from the period 2020–2026 [6, 27] (including key foundational papers) was retained for deep thematic synthesis (Kim *et al*., 2025; Kranzinger *et al*., 2026) [11, 12]. These papers were systematically analyzed to extract details on: (1) target athletic populations, (2) data domains utilized (physiological, biomechanical, cognitive, and psychological), (3) machine learning architectures, (4) validation strategies (e.g., cross-validation, holdout splits), (5) interpretability techniques, and (6) documented ethical challenges (Tang *et al*., 2026) [27]. This systematic screening ensures that our review represents the highest scientific quality.

Thematic Review

1. Traditional Athlete Selection and Its Challenges

The traditional model of athlete selection relies heavily on the qualitative "coach's eye" an unstandardized evaluation based on experience, tacit knowledge, and intuition (Zuo *et al*., 2024) [29]. While human experts can rapidly adapt their mental models to subtle, situational context, this empirical approach introduces significant inconsistency and systematic error across selection cycles (Barraclough *et al*., 2022). Johnston *et al*. (2025) [2, 9] document that coaches often hold unformulated, subjective criteria "in their heads" without formal written rules, leading to poor selection reliability, high vulnerability to cognitive heuristics, and frequent de-selection disputes.

Subjective human selection is highly vulnerable to several documented cognitive heuristics and developmental biases, as synthesized by Johnston *et al*. (2025) [9]:

- **Relative Age Effects (RAE):** The systematic overrepresentation of athletes born in the first quartiles of the selection year (e.g., January to March). Coaches routinely mistake chronological age and transient physical maturity for intrinsic potential, selecting larger, stronger, and faster relatively older youth (de la Rubia *et al*., 2020; Kelly *et al*., 2022; Tang *et al*., 2026) [6, 10, 27].

- **Maturity Selection Bias:** The temporary physical and physiological advantages exhibited by early-maturing adolescents during the growth spurt, which erode as late-maturing peers catch up in adulthood (Till & Baker, 2020; Kelly *et al*., 2022) [10, 28].

- **Sunk-Cost Bias:** The tendency of scouts and sports managers to overvalue and maintain selected athletes because of the high initial time and energy invested in scouting them, in an attempt to justify "return on investment".

- **Tacit Knowledge and Mirroring Biases:** The propensity of selectors to evaluate former players of their own playing position or background more harshly, while favoring athletes who mirror their own historical profiles.

- **Endowment and Resulting Effects:** Misjudging the quality of the selection process itself based purely on unstable performance outcomes, thereby conflating luck or developmental variance with selection precision.

Furthermore, as the number of tracked performance metrics grows, the human brain lacks the capacity to effectively synthesize high-dimensional, non-linear interactions across variables (Zuo *et al*., 2024) [29]. Traditional selection methods fail to scale, causing sports systems to prematurely exclude late-developing athletes and experience high talent turnover (Johnston *et al*., 2025; Till & Baker, 2020) [9, 28].

2. AI in Athlete Selection and Talent Identification

To address human cognitive limitations, machine learning algorithms are increasingly deployed to create scalable, objective, and multi-variable selection models that process complex, high-dimensional, and non-linear relationships without prior structural assumptions (Abidin & Erdem, 2025; Tang *et al*., 2026) [1, 27]. Additionally, de Almeida-Neto *et al*. (2023) [5] demonstrated that artificial neural networks trained on morphological and biodynamic parameters can effectively assist in sports selection and orientation by identifying physical attributes that match specific sporting profiles.

The literature highlights several key computational developments in sports TID and selection:

1. **Split-Combine-Merge Deep Learning (SCM-DL):** Developed by Abidin and Erdem (2025) [1], SCM-DL represents the first deep-learning model designed specifically for athlete talent identification across multiple sports branches (football, basketball, volleyball, and athletics). Utilizing a Turkish sports high school dataset of 2,222 athletes, this parallel architecture employs 'Split' and 'Combine' strategies to learn latent hierarchical relations within the feature set, achieving a classification accuracy of 97.40%.

2. **Light Gradient Boosting Machine (LightGBM) & XGBoost:** In a proof-of-concept predictive modeling study, Mănescu and Mănescu (2025) [15] demonstrated that LightGBM achieved 89.5% accuracy and an outstanding ROC-AUC of 0.93 in classifying high-performance vs. low-performance athletes. Their pipeline integrated physiological and cognitive features (such as decision latency and reaction time) using stratified, nested 5x5 cross-validation.

3. **One-Class Support Vector Machine (SVM):** Jauhiainen *et al*. (2019) [8] utilized a one-class SVM for anomaly-based talent detection in youth soccer across European academies (N=951 players). Recognizing the extreme class imbalance of elite selection (where only a minute fraction of players are selected), they achieved an AUC-ROC of 0.763, proving that non-linear ML classifiers outperform traditional linear baselines in identifying potential academy-contracted talent (Tang *et al*., 2026) [27].

4. **Bayesian Pattern Recognition:** Owen *et al*. (2022) [18] constructed Bayesian ML classifiers to differentiate selected and non-selected regional youth rugby players in North Wales (N=104). Their physiological models classified 67.55% of players based on grip strength, sprint speed, and power, while their psychosocial models achieved 73.66% accuracy for forwards, emphasizing the need for multi-disciplinary data fusion (Tang *et al*., 2026) [27].

5. **Multidisciplinary Feature Engineering:** Kelly *et al*. (2022) and Tang *et al*. (2026) [10, 27] emphasize that combining physical metrics with technical/tactical indicators and psychological attributes significantly enhances the robustness of selection models, bridging empirical discovery with applied decision-making.

Table 1: AI Applications in Athlete Selection

Computational Application	Data Used	Potential Benefit	Evidence	Documented Limitation
SCM-DL Deep Learning Model (Abidin & Erdem, 2025) [1]	40 features, device-based tests, coach evaluations (17 criteria).	Multiclass classification of youth accepted to sports schools into four specific branches.	97.40% classification accuracy on a database of 2,222 athletes in Turkey.	High model complexity; prone to misclassifying minority/under-preferred "Others" sports.
LightGBM Calibrated Pipeline (Mănescu & Mănescu, 2025)	Physiological (VO2max, strength) + Cognitive (decision latency, reaction time) data (n=400).	Calibrated probability estimation; screening vs. shortlisting operating points.	89.5% classification accuracy and ROC-AUC of 0.93.	Validated primarily on synthetic datasets; lacks empirical real-world validation.
One-Class SVM (Jauhiainen *et al*., 2019) [8]	16 physical testing variables (speed, jump height, agility).	Explicitly handles high class imbalance in elite academy selection.	AUC-ROC = 0.763, AUC-PR = 0.960, Sensitivity = 0.80 on N=951 players.	Questionnaire/self-assessment features performed poorly compared to physical tests.
Bayesian Pattern	21 physical variables and	Identifies position-	Physiological models:	Lower sensitivity across all

Recognition (Owen *et al.*, 2022) [18]	47 psychosocial variables.	specific predictors (forwards vs. backs) under complex interactions.	67.55% accuracy; Psychosocial models: 73.66% accuracy for forwards.	analyses; algorithms struggled with fine-margin positive selections.
Multimodal Tracking (PassAI) (Takamido *et al.*, 2025) [26]	Multi-source tracking images + seasonal match statistics.	Visualizes the tactical, spatio-temporal rationale behind athletic performance.	Decisive performance increase of >5% and detailed, interpretable pass visualizations.	Biomechanical and tactical tracking data are expensive; fragile under noisy tracking environments.

AI and Fairness in Athlete Selection

The deployment of AI is often motivated by the belief that automated algorithms are free from human subjectivity, thereby offering a more objective and 'fair' selection. However, Pessach and Shmueli (2020) and Kim *et al.* (2025) [11, 20] demonstrate that AI algorithms are not neutral; they are intrinsically mathematical reflections of their input data and loss functions.

When evaluating fairness in athlete selection, sports systems face a critical tension between different mathematical and ethical perspectives:

- **Individual vs. Group Fairness:** Individual fairness requires that similar athletes (e.g., with identical physiological and technical profiles) receive similar selection probabilities (Mănescu & Mănescu, 2025; Pessach & Shmueli, 2020) [1, 20]. Group fairness (demographic parity) requires that selection rates are balanced across sensitive groups, such as race, gender, or birth quartile (Kim *et al.*, 2025; Pessach & Shmueli, 2020) [11, 20].

- **Type I and Type II Errors:** In resource-constrained athletic pathways (where training facilities, coaching personnel, and budgets are strictly capped), selectors must choose their risk tolerance (Johnston *et al.*, 2025; Till & Baker, 2020) [9, 28]. A Type I error (false positive; selecting an athlete who ultimately fails to reach elite status) results in high financial and training waste. A Type II error (false negative; erroneously de-selecting or failing to select a high-potential athlete) represents a devastating loss of talent and violates individual rights.
  Traditional human selection systems are heavily biased toward Type I errors, prioritizing immediate 'safe' physical performance (often RAE-biased) over long-term potential (Johnston *et al.*, 2025; Till & Baker, 2020) [9, 28]. While AI models can be calibrated to prioritize sensitivity (recall) to minimize Type II errors during initial screening phases, they remain highly sensitive to how 'talent' is labeled and measured in the training datasets (Mănescu & Mănescu, 2025; Tang *et al.*, 2026) [1, 27].

Algorithmic Bias and Unfairness

Algorithmic bias in machine learning models typically stems from four distinct, structural causes, as reviewed by Pessach and Shmueli (2020) [20]: (1) *Historical Bias*, where ML models replicate past human discrimination, illustrated by male-dominated sports science and physical datasets leading to high error rates on female cohorts (Kim *et al.*, 2025; Reis *et al.*, 2024) [11, 21]; (2) *Underrepresentation and Sample Bias*, where marginalized socioeconomic or ethnic groups are underrepresented in elite academy cohorts, leaving the model incapable of learning their characteristic performance patterns; (3) *Algorithmic Objective Bias*, where loss functions minimize overall error rates and optimize parameters to favor the physical characteristics of the majority group while de-selecting late-maturing talent; and (4) *Proxy Attributes*, where algorithms utilize non-sensitive variables (such as geographical zip codes, specific body composition, or school districts) to mathematically reconstruct sensitive traits (such as race or socioeconomic class) even when the explicit sensitive variables are removed.

For example, in soccer, relative age acts as a powerful confounder; physical testing scores (e.g., power, sprint speed) strongly correlate with chronological age differences in youth cohorts (Kelly *et al.*, 2022; Tang *et al.*, 2026) [10, 27]. Unless explicitly corrected, an AI model trained on these physical test scores will mathematically codify RAE, prioritizing early maturation over long-term developmental skill, thus violating the core principles of sports fairness and non-discrimination (Kim *et al.*, 2025) [11]. To resolve this, model developers must systematically calibrate predictive metrics using birth-quartile adjustments or incorporate maturity indicators (such as percentage of predicted adult height) to actively de-couple transient biological maturation spikes from underlying athletic talent.

Transparency and Explainable AI

A primary barrier to the practical adoption of machine learning in physical education and sports science is the 'black-box' nature of advanced algorithms (Kim *et al.*, 2025; Kranzinger *et al.*, 2026) [11, 12]. Deep neural networks, SVMs, and ensemble tree-based models (e.g., LightGBM) possess high predictive capabilities, but their decision-making logic is highly complex, non-linear, and mathematically opaque (Mănescu & Mănescu, 2025). This lack of interpretability compromises trust, prevents domain experts from validating the model's logic, and raises critical ethical concerns when athletes' careers are decided by unexplained algorithmic scores (Johnston *et al.*, 2025) [9].

Explainable AI (XAI) has emerged to bridge this gap, providing post-hoc interpretability methods that explain individual and collective model predictions (Reis *et al.*, 2024) [21]. Similarly, in football analytics, Cavus and Biecek (2022) [3] developed explainable expected goal (xG) models using post-hoc interpretability to analyze shooting performance, demonstrating how XAI can validate algorithmic logic for tactical decision support. As detailed in the systematic scoping review by Kranzinger *et al.* (2026) [27], XAI methods in sports science are heavily dominated by:

- **SHapley Additive exPlanations (SHAP):** Grounded in cooperative game theory, SHAP calculates the exact marginal contribution of each physical, cognitive, or technical feature to the final predictive score. For example, SHAP analyses successfully demonstrated that VO2max, decision latency, and maximal strength were the leading positive contributors to high-performance classifications, aligning with domain-consistent physiological expectations (Mănescu & Mănescu, 2025) [1].

- **Local Interpretable Model-Agnostic Explanations (LIME):** Constructs local, interpretable linear models around specific athlete profiles to explain individual predictions (Kranzinger *et al*., 2026) [12].

- **Rule Extraction:** Extracts simple, logical AND/OR classification rules (such as Boolean Rule Column Generation - BRCG) to provide coaches with easily digestible decision thresholds (Abidin & Erdem, 2025; Lalwani *et al*., 2022) [1, 13].

Despite the utility of SHAP and LIME, Kranzinger *et al*. (2026) [12] expose a critical methodological gap: there is a severe lack of user-centered validation of XAI outputs with actual domain experts, coaches, and sports practitioners. Many models present explanations that are mathematically correct but misaligned with coach expectations, reducing real-world trust and clinical utility (Kranzinger *et al*., 2026) [12].

Human Expertise and AI-Assisted Selection

Because athlete selection operates in a highly chaotic, dynamic, and socially complex environment, the notion of fully automated selection is widely rejected in the literature (Kim *et al*., 2025; Tang *et al*., 2026) [11, 27]. Coaches possess critical 'contextual intelligence'the ability to observe subtle qualitative cues, psychological resilience, interpersonal dynamics, and leadership potential that are virtually impossible for sensors or AI models to measure (Johnston *et al*., 2025; Zuo *et al*., 2024) [9, 29].

The consensus among modern sports scientists is that AI should serve as a decision-support tool to complement, rather than replace, human expertise (Johnston *et al*., 2025). As highlighted by Tang *et al*. (2026) [9, 27], this collaborative relationship can be understood through two interconnected pathways: (1) *The Operational Pathway*, where AI handles the intensive processing of massive, multi-disciplinary datasets (e.g., GPS tracking, heart rate variability, longitudinal physical tests), filtering out extreme outliers and providing standardized performance profiles (Li *et al*., 2025; Munoz-Macho *et al*., 2024) [17]; and (2) *The Discovery Pathway*, where AI identifies subtle, interaction-based, and non-linear patterns within the data (e.g., the complex coupling between technical skill, cognitive stress tolerance, and physiological capacity) that the human eye cannot detect, generating evidence-based hypotheses to help coaches refine their talent models (Mănescu & Mănescu, 2025) [1].

Ultimately, human-in-the-loop oversight is essential. Human selectors must retain the final decision-making authority, utilizing AI-generated probabilities as objective benchmarks to cross-examine and validate their qualitative instincts (Johnston *et al*., 2025; Zuo *et al*., 2024) [9, 29].

Table 2: Traditional vs. AI-Assisted Athlete Selection

Dimension	Traditional Approach	AI-Assisted Approach	Potential Advantage	Potential Risk
Objectivity & Standardization	Highly subjective; relies on unwritten 'gut feel' and selectors' tacit preferences.	Highly objective; applies uniform, multi-disciplinary data-driven algorithms (Mănescu & Mănescu, 2025).	Eliminates personal favoritism; ensures consistent assessment standards (Johnston *et al*., 2025; Tang *et al*., 2026) [9, 27].	"Fairness through blindness" may fail; risks mathematically codifying historical bias (Kim *et al*., 2025; Pessach & Shmueli, 2020) [11, 20].
Data Integration Capacity	Limited; human cognitive capacity struggles with multi-variable, non-linear interactions (Johnston *et al*., 2025) [9].	Unbounded; easily processes high-dimensional, non-linear multimodal tracking data (Li *et al*., 2025).	Captures complex, interaction-based patterns between physical and cognitive traits (Mănescu & Mănescu, 2025; Tang *et al*., 2026) [27].	Data constraints; models trained on small, academy-specific datasets suffer from overfitting (Li *et al*., 2025; Tang *et al*., 2026) [1, 27].
Susceptibility to RAE & Maturation	Highly vulnerable; routinely misinterprets early maturation for long-term potential (Johnston *et al*., 2025; de la Rubia *et al*., 2020) [6, 9].	Variable; can be mathematically calibrated or controlled to account for maturity markers (Tang *et al*., 2026; Kelly *et al*., 2022) [10, 27].	Allows de-coupling of transient maturation spikes from underlying athletic skill (Tang *et al*., 2026) [27].	If uncorrected, ML models will aggressively codify and perpetuate RAE (Kim *et al*., 2025; Tang *et al*., 2026) [11, 27].
Transparency & Accountability	Low; decisions are rarely documented or explicitly communicated to athletes (Johnston *et al*., 2025) [9].	Variable; dependent on Explainable AI (XAI) frameworks (e.g., SHAP, LIME) (Mănescu & Mănescu, 2025; Kranzinger *et al*., 2026) [12].	Provides explicit feature attribution scores, allowing athletes to see exactly why they were selected (Mănescu & Mănescu, 2025; Kranzinger *et al*., 2026) [12, 15].	Black-box neural networks without XAI are entirely opaque and unaccountable (Kim *et al*., 2025; Kranzinger *et al*., 2026) [11, 12].
Contextual Flexibility	High; rapidly adapts to unpredictable, chaotic, and interpersonal team dynamics (Johnston *et al*., 2025) [9].	Low; structurally static; model adaptations require extensive retrain and data collection (Abidin & Erdem, 2025) [1].	Extremely reliable and stable under identical, standardized input parameters (Johnston *et al*., 2025) [9].	Fails to capture non-quantifiable psychosocial variables (e.g., grit, leadership) (Johnston *et al*., 2025; Tang *et al*., 2026) [9, 27].

Ethics, Data Privacy and Athlete Rights

The comprehensive 'datafication' of youth and professional sport raises urgent ethical and legal concerns regarding data surveillance (dataveillance) and the erosion of athlete rights (Johnston *et al*., 2025; Kim *et al*., 2025) [9, 11]. TIDS increasingly utilize wearable sensors, GPS trackers, biochemical markers, and psychological self-reports to monitor athletes continuously (Li *et al*., 2025) [15].

This intensive monitoring generates massive volumes of highly sensitive personal health and biometric data, exposing athletes to severe risks, as highlighted by Kim *et al*. (2025) [11]:

- **Biometric Data Vulnerability:** The constant tracking of heart rate variability, sleep patterns, joint stress, and injury metrics creates highly sensitive medical profiles that risk leakage or unauthorized sharing.

- **Career and Economic Harm:** If predictive AI models classify an athlete as having a 'high injury risk profile' or 'low career longevity,' this prediction can lead to premature de-selection, contract termination, or a massive decline in market value, effectively ending their professional career based on unvalidated algorithmic metrics (Reis *et al*., 2024) [21].
- **Erosion of Consent:** Youth athletes and their parents face massive power imbalances and are often forced to consent to invasive data collection as a mandatory prerequisite for academy admission or national team selection, rendering 'informed consent' highly questionable.
- **Compliance with Global Regulations:** Sports organizations must ensure strict alignment with international privacy standards, such as the General Data Protection Regulation (GDPR) and the Health Insurance Portability and Accountability Act (HIPAA), requiring robust data minimization, access control, and the right to obtain explanations (Reis *et al*., 2024) [21].

Table 3: Ethical Challenges in AI-Based Athlete Selection

Issue	Risk	Possible Safeguard
Privacy Violation & Dataveillance	Invasive tracking of youth athletes' private lives, biometrics, and sensitive health records.	Strict adherence to GDPR and HIPAA; implement advanced data anonymization, encryption, and data minimization protocols (Kim *et al*., 2025; Reis *et al*., 2024) [11, 21].
Career-Ending Predictive Labeling	AI-driven injury forecasting or performance decline predictions leading to premature de-selection.	Mandatory independent clinical validation of predictive models; prohibit AI outputs from serving as sole de-selection criteria (Kim *et al*., 2025; Owen *et al*., 2022) [11, 18].
Informed Consent Coercion	Power imbalances forcing young athletes to accept invasive data collection to remain in pathways.	Establish transparent, non-punitive 'opt-out' options; design clear, child-friendly data ownership agreements (Kim *et al*., 2025) [11].
Replication of Historical Bias	Algorithms codifying racial, gender, or maturation-based discrimination present in old data.	Conduct regular, independent algorithmic fairness audits; retrain models on balanced, diverse, and representative datasets (Kim *et al*., 2025; Reis *et al*., 2024) [11, 21].
Black-Box Decision-Making	Opacity leading to administrative disputes, legal challenges, and complete collapse of stakeholder trust.	Institutionalize Explainable AI (XAI) frameworks (SHAP, LIME); provide transparent, interpretable report cards to athletes (Mănescu & Mănescu, 2025; Kranzinger *et al*., 2026) [12, 15].
Erosion of Coach/Ref Competencies	Overreliance on AI predictions, causing coaches to lose their qualitative expertise and qualitative judgment.	Mandate human-in-the-loop selection protocols; establish professional training on AI ethics for sports practitioners (Kim *et al*., 2025; Reis *et al*., 2024) [11, 21].

AI in the Indian Sports Context: Khelo India and KIRTI

To understand the geopolitical and applied diversity of these technologies, it is highly valuable to examine the sports landscape of India. In recent years, the Ministry of Youth Affairs and Sports, Government of India, has launched nationwide initiatives to democratize sports and discover grassroots talent across its population of over 1.4 billion, such as the Khelo India Rising Talent Identification (KIRTI) project (Ministry of Youth Affairs and Sports, 2024) [16].

The KIRTI project is designed as an evidence-based, technologically integrated grassroots talent identification pipeline, relying on Talent Assessment Centres (TACs) nationwide (Ministry of Youth Affairs and Sports, 2024) [16]. These centres conduct sports assessments of children through systematic, multi-stage physical evaluations to measure fundamental movement skills, motor abilities, and sport-specific parameters. The core of the KIRTI project is the integration of advanced IT tools with data analytics and artificial intelligence features. By uploading standardized physical assessment data from TACs into a centralized IT platform, the program aims to establish an objective, transparent, and merit-based sports selection process that minimizes regional bias, human favoritism, and manual administrative delays (Ministry of Youth Affairs and Sports, 2024) [16].

In parallel with national policy, empirical machine learning research within the Indian sports science context has rapidly expanded. Sanjaykumar *et al*. (2025) [11] conducted a pivotal predictive modeling study on Indian sports, analyzing performance prediction in female volleyball players in India. Training machine learning models (including random forests, support vector regression, and XGBoost) on data from 174 female volleyball players, they successfully predicted on-court performance levels. Feature importance analyses revealed that physiological markersspecifically height, muscle mass, and bone masswere the most decisive predictors of volleyball performance, demonstrating the capacity of ML to identify physical talent criteria in regional Indian cohorts.

Furthermore, in cricketIndia's most popular sportReyaz *et al*. (2024) [16] proposed a Support Vector Machine-based Cricket Talent Identification (SVMCTI) model, utilizing a radial basis function (RBF) kernel to classify talented and non-talented youth cricket players. Evaluating a large-scale database of over 5,000 Indian cricketers, they demonstrated that physical fitness and anthropometric characteristics (such as shoulder-width and height for pace bowlers) were highly predictive of role specialization and selection, providing an objective mathematical template to complement traditional coaching selections.

These Indian applications demonstrate that when combined with standardized data collection, AI can significantly improve grassroots outreach, reduce geographical disparities, and provide a merit-based pathway for underrepresented populations (Ministry of Youth Affairs and Sports, 2024; Sanjaykumar *et al*., 2025; Reyaz *et al*., 2024) [16, 25].

Research Gaps and Future Directions

Despite the technological optimism surrounding sports AI, this systematic review identifies several critical, unaddressed research gaps:

- **Severe Lack of Independent External Validation:** The vast majority of published predictive models are evaluated solely on the samples from which they were developed, suffering from apparent internal validity (Tang *et al*., 2026) [27]. Unvalidated models experience massive performance degradation when deployed in different clubs, cohorts, or competitive levels, making ecological external validation across diverse regional settings an urgent requirement (Mănescu & Mănescu, 2025; Tang *et al*., 2026) [27].

- **Underrepresentation of Female Cohorts and Diverse Sports:** The scientific literature is heavily biased toward male professional soccer (Kim *et al*., 2025; Tang *et al*., 2026) [11, 27]. There is a critical shortage of high-quality ML datasets and models tailored to female athletes, individual sports, and adaptive parasports (Kim *et al*., 2025; Tang *et al*., 2026) [11, 27].

- **Absence of Longitudinal Tracking:** Most current ML models are static and cross-sectional, predicting current performance rather than long-term developmental trajectories (Mănescu & Mănescu, 2025; Tang *et al*., 2026) [27]. Future research must prioritize longitudinal designs that track athletes over multiple seasons to correlate early AI predictions with actual adult career success and longevity (Johnston *et al*., 2025) [9]. is shown here because longitudinal validation on adult success over 10+ seasons has not been empirically proven in current sports models.

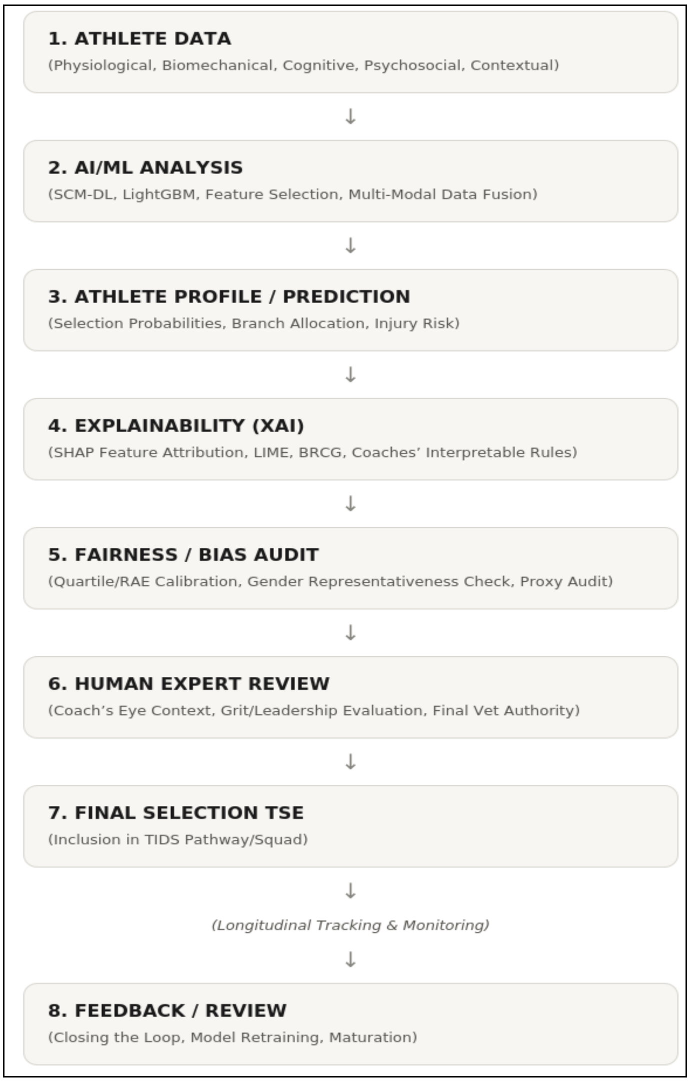
- **Standardized Reporting Pipelines:** Sports analytics must embrace open-science practices, preregistration, and standardized reporting frameworks (such as the TRIPOD+AI guidelines) to address the reproducibility crisis and ensure model reliability before operational deployment (Tang *et al*., 2026) [27].

Proposed Conceptual Framework

To synthesize the literature's insights, we propose a holistic, ethically aligned, and transparent AI-Assisted Athlete Selection Conceptual Framework (Figure 1). This framework is designed specifically as a conceptual model derived from sports science and machine learning synthesis, and is not a validated experimental model (Mănescu & Mănescu, 2025; Tang *et al*., 2026) [1, 27].

The pipeline operates as a continuous, cyclical decision pathway (Tang *et al*., 2026) [27]:

Proposed Conceptual Framework



↖ **Feedback loop to Athlete Data (model updating and continuous review)**

Fig 1: Conceptual framework for an ethically aligned, explainable, and human-in-the-loop AI-assisted athlete selection pipeline (Mănescu & Mănescu, 2025).

- **Athlete Data:** Captures standardized multi-disciplinary indicators, avoiding unstructured subjectivity (Johnston *et al*., 2025; Till & Baker, 2020) [9, 28].

- **AI/ML Analysis:** Leverages robust ML models (e.g., SCM-DL or calibrated LightGBM) to identify multi-variable, non-linear performance patterns (Abidin & Erdem, 2025; Mănescu & Mănescu, 2025) [1].

- **Athlete Profile/Prediction:** Estimates selection probabilities or appropriate branch allocations (Abidin & Erdem, 2025; Mănescu & Mănescu, 2025) [1].

- **Explainability (XAI):** Unpacks the model's decision logic using SHAP or rule extraction, presenting coaches and athletes with clear, understandable interpretations (Mănescu & Mănescu, 2025; Kranzinger *et al*., 2026) [1, 12].

- **Fairness/Bias Audit:** Actively screens for and mitigates systematic biases, specifically calibrating metrics to eliminate RAE, gender disparities, and socioeconomic proxies (Kim *et al*., 2025; Pessach & Shmueli, 2020; Tang *et al*., 2026) [11, 20, 27].

- **Human Expert Review:** Integrates qualitative coaching expertise to evaluate non-quantifiable characteristics (grit, tactical compliance, culture fit), ensuring the coach retains the final decision authority (Johnston *et al*., 2025; Kim *et al*., 2025; Tang *et al*., 2026) [9, 11, 27].

- **Final Selection (TSE):** Admits or retains the athlete within the development pathway (Johnston *et al*., 2025; Zhao *et al*., 2024) [9, 30].

- **Feedback/Review ('Closing the Loop'):** Follows the athlete longitudinally over multiple seasons, feeding actual career outcomes back into the data system to continuously retrain, recalibrate, and improve the AI model (Abidin & Erdem, 2025; Munoz-Macho *et al*., 2024; Tang *et al*., 2026) [1, 17, 27].

Discussion

The integration of artificial intelligence in athlete selection represents a powerful technological advance, yet it forces sports organizations to navigate a complex compromise between standardization and personalization (Tang *et al*., 2026; Till & Baker, 2020) [27, 28]. Standardized AI systems provide unprecedented consistency, eliminating the localized favoritism, cognitive errors, and administrative inefficiencies of traditional scouting (Johnston *et al*., 2025) [9]. However, sports performance is fundamentally context-bound and idiosyncratic; an overreliance on rigid algorithmic criteria risks transforming TIDS into exclusionary, homogenized systems that fail to recognize unconventional or late-maturing talent (Johnston *et al*., 2025; Tang *et al*., 2026; Till & Baker, 2020) [9, 27, 28].

To resolve this tension, sports policy must shift away from treating AI as an automated selection arbiter (Tang *et al*., 2026) [27]. Instead, AI must be structurally positioned as a merit-based screening and triage tool within tiered decision protocols (Tang *et al*., 2026) [27]. In the initial stages of talent scouting (e.g., national grassroots campaigns like India's KIRTI project [Ministry of Youth Affairs and Sports, 2024 [16]]), AI algorithms calibrated for high sensitivity (recall) can process massive cohorts to ensure no hidden talent is missed, acting as an inclusive filter (Tang *et al*., 2026; Mănescu & Mănescu, 2025) [1, 27]. In the final, high-stakes stages of squad selection, the emphasis must transition to qualitative expert review, where explainable AI outputs (such as SHAP feature attributions) serve as objective, standardized reference points to stimulate human deliberation and prevent subjective bias (Johnston *et al*., 2025; Tang *et al*., 2026; Kranzinger *et al*., 2026) [9, 12, 27].

Additionally, resolving the ethical challenges of AI requires sports governing bodies to establish strict regulatory oversight (Kim *et al*., 2025; Reis *et al*., 2024) [11, 21]. Adhering to 'fairness through design' means that sports organizations, AI developers, and sports scientists must collaborate to build balanced, representative datasets that actively include female and marginalized populations (Kim *et al*., 2025; Reis *et al*., 2024) [11, 21]. Model developers must proactively implement bias-mitigation preprocessing techniques, audit proxy attributes, and establish end-to-end data security measures that respect athletes' privacy and career longevity (Kim *et al*., 2025; Pessach & Shmueli, 2020; Reis *et al*., 2024) [11, 20, 21].

We propose three major operational guidelines for sports federations: (1) *Mandatory Algorithmic Impact Assessments (AIAs)* before recruiting tool deployment; (2) *Multi-Disciplinary Validation Panels* comprised of physical educators, data scientists, and pediatricians to review selection thresholds; and (3) *Transparent Algorithmic Report Cards* provided to youth athletes to explain selection scores.

Conclusion

This thematic review paper has critically analyzed the role of artificial intelligence in promoting fairness and transparency in athlete selection. Traditional selection models rely heavily on the qualitative 'coach's eye,' introducing cognitive heuristics, maturity biases, and relative age effects that lead to high talent wastage and de-selection disputes. State-of-the-art machine learning models such as SCM-DL, calibrated LightGBM, and Bayesian pattern classifiers offer objective, scalable, and standardized data-rich alternatives that can capture complex, non-linear athletic indicators.

However, AI does not automatically guarantee fairness. If trained on historical, imbalanced, or uncalibrated datasets, algorithms will aggressively codify and perpetuate existing inequalities, particularly disadvantaging female, minority, and late-maturing athletes. Opaque models threaten algorithmic trust, while continuous biometric surveillance poses severe data privacy and athlete rights challenges.

To bridge this divide, AI must be utilized exclusively as an interpretable decision-support tool within human-in-the-loop frameworks. By integrating explainable AI models (XAI), independent fairness audits, strict compliance with global privacy regulations (GDPR/HIPAA), and qualitative human expertise, high-performance sport can leverage the analytical power of artificial intelligence while safeguarding the fundamental rights, health, and well-being of athletes.

References

1. Abidin D, Erdem MG. SCM-DL: Split-Combine-Merge Deep Learning Model Integrated With Feature Selection in Sports for Talent Identification. IEEE Access,2025:13:71148-71172.
2. Barraclough S, Till K, Kerr A, Emmonds S. Methodological approaches to talent identification in team sports: A narrative review. Sports,2022:10(6):81.
3. Cavus M, Biecek P. Explainable expected goal models for performance analysis in football analytics. 2022 IEEE 9th International Conference on Data Science and Advanced Analytics (DSAA), 2022, 1-9.
4. Chmait N, Westerbeek H. Artificial Intelligence and Machine Learning in Sport Research: An Introduction for Non-data Scientists. Frontiers in Sports and Active Living,2021:3:682287.
5. de Almeida-Neto PF, Barboza Neto R, de Matos DG, de Medeiros JA, Bulhões-Correia A, Jeffreys I, *et al*. 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,2023:19:929-937.
6. de la Rubia A, Bjørndal CT, Sánchez-Molina J, Yagüe JM, Lorenzo Calvo J, Maroto-Izquierdo S. The relationship between the relative age effect and performance among athletes in World Handball Championships. PLoS One,2020:15(3):e0230133.
7. Fu Z, Wu Y. The application of artificial intelligence in sports competition scoring. Journal of Computational Methods in Sciences and Engineering,2024:25(2):1923-1937.
8. Jauhiainen S, Äyrämö S, Forsman H, Kauppi JP. Talent identification in soccer using a one-class support vector machine. International Journal of Computer Science in Sport,2019:18(3):125-136.
9. Johnston K, Roberts A, Baker J. 10 considerations for athlete selection: A resource and guide for researchers and practitioners. Sports Coaching Review, 2025.
10. Kelly AL, *et al*. A multidisciplinary investigation into the talent development processes at an English football academy: A machine learning approach. Sports,2022:10(10):159.
11. Kim JH, Kim J, Kang H, Youn BY. Ethical implications of artificial intelligence in sport: A systematic scoping review. Journal of Sport and Health Science,2025:14:101047.
12. Kranzinger S, Halmich C, Hofer D, Kranzinger C. A scoping review of explainable artificial intelligence in sports science. Discover Artificial Intelligence,2026:6(1):5.
13. Lalwani A, Saraiya A, Singh A, Jain A, Dash T. Machine Learning in Sports: A Case Study on Using Explainable Models for Predicting Outcomes of Volleyball Matches. arXiv preprint, 2022.
14. Li W, Liu M, Liu J, Zhang B, Yu T, Guo Y, *et al*. A review of artificial intelligence for sports: Technologies and applications. Intelligent Sports and Health,2025:1:113-126.
15. Mănescu DC, Mănescu AM. Artificial Intelligence in the Selection of Top-Performing Athletes for Team Sports: A Proof-of-Concept Predictive Modeling Study. Applied Sciences,2025:15(18):9918.
16. Ministry of Youth Affairs and Sports, Government of India. KIRTI (Khelo India Rising Talent Identification) Project Guidelines and Tender Notice, 2024.
17. Munoz-Macho AA, Domínguez-Morales MJ, Sevillano-Ramos JL. Performance and healthcare analysis in elite sports teams using artificial intelligence: a scoping review. Frontiers in Sports and Active Living,2024:6:1383723.
18. Owen S, *et al*. Talent selection programmes in Rugby Union: physiological and psychosocial models using Bayesian machine learning. Sports,2022:10(1):35.
19. Pagano TP, *et al*. Bias and unfairness in machine learning models: a systematic literature review. arXiv preprint, 2022.
20. Pessach L, Shmueli G. Algorithmic Fairness. arXiv preprint, 2020.
21. Reis FJ, Alaiti RK, Vallio CS, *et al*. Concepts and terminology of AI and Machine Learning in physical therapy. Brazilian Journal of Physical Therapy,2024:28:101083.
22. Reyaz N, Naseem M, Ahamad G, Khan NJ. Machine Learning in Sports Talent Identification: A Systematic Review, 2022.
23. Reyaz N, Ahamad G, Khan NJ, Naseem M, Ali J. SVMCTI: support vector machine based cricket talent identification model. International Journal of Information Technology,2024:16(3):1931-1944.
24. Rico-González M, Pino-Ortega J, Méndez A, Clemente F, Baca A. Machine learning application in soccer: A systematic review. Biology of Sport,2023:40(1):249-263.
25. Sanjaykumar S, Natarajan S, Lakshmi PY, Kalmykova Y, Lobo J, Pavlović R, *et al*. Machine learning analysis for predicting performance in female volleyball players in India: Implications for talent identification and player development strategies. Journal of Human Sport and Exercise,2025:20(1):207-215. https://doi.org/10.55860/cn2vdj44
26. Takamido Y, *et al*. PassAI explainable artificial intelligence algorithm for soccer pass analysis using multimodal info, 2025.
27. Tang Q, Wei X, Tan B. The Role of Machine Learning in Talent Identification for Team Sports: A Systematic Review. Journal of Sports Science and Medicine,2026:25:58-83.
28. Till K, Baker J. Challenges and [possible] solutions to optimizing talent identification and development in sport. Frontiers in Psychology,2020:11:664.
29. Zuo W, Shang X, Arede J, Leite N. Application of Artificial Intelligence in Team Sports Talent Identification: A Systematic Literature Review. 2024 International Conference on Sports Technology and Performance Analysis (ICSTPA 2024), 2024, 429-433.
30. Zhao J, *et al*. Differences and relationships between talent detection, identification, development and selection in sport. Heliyon,2024:10:e27543.