

**International Journal of
PHYSICAL EDUCATION, FITNESS AND SPORTS**

Electronic Supplementary Information

**Artificial Intelligence in Sports Science: Evidence
Maturity, Athlete Digital Twins, Multimodal Analytics,
and a Translation Agenda for 2026–2035**

S. Akila ^{a, *}

^a Department of Physical Education and Sports, Central University of Tamil Nadu, Thiruvavur-610005, Tamil Nadu, India

*Corresponding Author E-mail: akilas@cutn.ac.in

DOI: <https://doi.org/10.54392/ijpefs2633>

Received: 25-05-2026; Revised: 09-08-2026; Accepted: 17-08-2026; Published: 29-08-2026

Appendix A.

Table S1. Database-specific search documentation

Database	Search concept structure
Scopus	AI/ML/computer vision/foundation model/generative AI/wearable/multimodal/digital twin AND sport/athlete/physical education/sports medicine/exercise AND performance/injury/rehabilitation/biomechanics/coaching/monitoring/decision support
Web of Science Core Collection	Same three-concept structure adapted to Topic fields
PubMed/MEDLINE	Free-text and MeSH-compatible terms for AI, sport/exercise, and application domains
SPORTDiscus	Subject terms and free-text sport/AI combinations
IEEE Xplore	Metadata/abstract terms for AI technologies and sports applications
ACM Digital Library	Title/abstract terms for AI technologies and sports applications

Table S2. Full article-level extraction matrix for the 32 included studies

ID	Study	Category	Population /corpus	Sample	Sport/setting	Data modality	AI method/ model	Validation/ review method	Metric	Main finding	External validation	Calibration	Explainability	Fairness	Implementation	Maturity
1	Claudino (2019)	Systematic review	58 included studies; pooled sample 6,456 athletes (97% male; 76% professional)	58 studies; 6,456 athletes	Team sports	Structured performance and injury-risk data across 12 team sports	11 AI techniques ; most frequent: artificial neural networks, decision trees, support vector machines and Markov processes	Systematic review of PubMed, Scopus and Web of Science	Study-level predictive/classification metrics	AI was most frequently applied in soccer, basketball, handball and volleyball; prospective evaluation was identified as a major need.	Not applicable to review; prospective /external validation was limited in the reviewed literature	Rarely reported in reviewed studies	Model interpretability not systematically established	Pooled evidence was strongly male dominated (97%)	Evidence synthesis; predominantly retrospective technical studies	Contextual evidence synthesis
2	Reis (2024)	Review	Narrative/masterclass synthesis of published sports-AI literature	NR (narrative review)	Sports science	Performance, injury, rehabilitation and sport-management data	Machine learning, deep learning and related AI methods	Narrative conceptual review	Conceptual and application-level synthesis	AI has broad potential in sports, but implementation is constrained by data quality, interpretability, ethics and limited real-world validation.	Not applicable to review	Identified as an underdeveloped reporting area	Explainability identified as a central challenge	Bias and representativeness discussed conceptually	Evidence-synthesis level	Contextual evidence synthesis
3	Souaifi (2025)	Scoping review	Published studies on wearable technology, motion analysis and injury	NR in accessible abstract	Sports biomechanics	Wearable, kinematic, kinetic, temporal and motion-	Supervised ML, deep learning and multimodal analytical approaches	Scoping review	Application distribution, model performance and methodological	AI supports biomechanical pattern recognition and field-based analysis,	Limited across the reviewed literature	Infrequently reported	Interpretability identified as a persistent	Population representation inconsistently reported	Evidence synthesis with predominantly technical validation	Contextual evidence synthesis

			prevention in sports biomechanics			analysis data			characteristics	but temporal validation, interpretability, standardisation and generalisability remain limited.			limitation			
4	Van Eetvelde (2021)	Systematic review	11 original studies selected from 249 records	11 studies	Sport injury	Sport-injury risk factors and athlete monitoring variables	Multiple ML classifiers	Systematic review; Newcastle–Ottawa Scale and GRADE appraisal	Discrimination/classification performance and risk-of-bias assessment	ML can identify high-risk athletes and influential risk factors, but the evidence was limited by methodological heterogeneity and risk of bias.	Limited; most studies used internal validation	Generally not reported	Variable and usually limited	Small and heterogeneous athlete samples	Retrospective predictive evidence	Contextual evidence synthesis
5	Jauhiainen (2022)	Case-control predictive modelling	791 female elite handball and soccer players; 60 ACL injuries (title refers to an 880-athlete screening cohort)	791 analysed; 60 ACL injuries	Female athletes	3D motion analysis and physical screening-test variables	Four classifiers; linear support vector machine performed best	100 repeated cross-validation runs with permutation testing and class-imbalance analyses	Mean ROC-AUC; best model approximately 0.63	The extensive screening battery had statistically detectable but clinically insufficient ACL-injury predictive ability.	No independent external cohort	Not reported	Risk-factor contribution examined, but no full clinical explanation framework	Female elite handball and soccer athletes only	Retrospective internal validation	Level 2: retrospective internal validation
6	Seshadri (2021)	Narrative review	Published evidence on wearable monitoring of athlete workload	NR (narrative review)	Athletes	Physiological, workload, cardiovascular and movement	Wearable analytics, ML and AI-supported monitoring	Narrative review	Clinical and performance utility across use cases	Wearables can complement workload management and	Not applicable to review	Not systematically evaluated	Human interpretation remains	Representation not systematically appraised	Evidence-synthesis level	Contextual evidence synthesis

			and injury burden			t data from wearable sensors				return-to-play decisions, but validity, integration and interpretation require multidisciplinary oversight.			necessarily			
7	Chidambaram (2022)	Review	Original studies using sport sensors followed by AI processing for diagnostic, treatment or monitoring purposes	NR in accessible abstract	Sports medicine	Physiological, positional and kinematic wearable-sensor data	Multiple AI techniques applied to sensor data	Narrative review; MEDLINE, Embase and Scopus search	Diagnostic, monitoring and performance-optimisation outcomes	AI-enhanced sensing can detect meaningful physiological and movement patterns, although clinical validation and reporting remain heterogeneous.	Limited across included studies	Not systematically reported	Model transparency inconsistently reported	Sport and participant representation heterogeneous	Evidence synthesis; mostly technical/observational applications	Contextual evidence synthesis
8	Gámez Díaz (2020)	Survey/review	Published digital-twin and virtual-coaching systems for physical activity	NR in accessible abstract	Physical activity	Multimodal physical-activity, sensor and interaction data	Machine learning and digital-twin coaching architectures	Survey and taxonomy	Interactivity, objectives, privacy/security and system characteristics	The paper defined digital-twin coaching and identified multimodal interaction, privacy, security and standardisation as central developments	Not applicable to survey	Not applicable	System transparency discussed conceptually	User inclusion not systematically quantified	Conceptual and prototype systems	Level 0–1: conceptual/prototype evidence

										nt priorities.						
9	Fister Jr. (2021)	Conceptual/technical article	Cycling artificial-sport-trainer prototype implemented on a Raspberry Pi	Prototype demonstration; athlete sample NR	Sport training	Cycling training and sensor inputs	Computational-intelligence artificial sport trainer with cognitive digital-twin framework	Technical prototype and platform porting	Feasibility of realisation and control on embedded hardware	The digital-twin framework was technically feasible on low-cost embedded hardware, but no prospective athlete-outcome evaluation was demonstrated.	No	Not applicable	Conceptual cognitive model described	Not assessed	Working prototype	Level 1: prototype/technical feasibility
10	Boillet (2024)	Empirical modelling study	National-level cyclists used to create individual physiology-based models	Small cyclist validation cohort; exact n NR in accessible abstract	Cycling	Physiological characteristics, field performance and energetic simulation data	Individualised physiology-based digital twin; reinterpreted Margaria-Morton model	Comparison with field performance, experimental observations and published physiological data	Simulated record times, 3-min all-out tests, metabolites and oxygen kinetics	Individualised twins reproduced field and experimental behaviour sufficiently to support athlete-specific performance simulation.	Validation used field observations but not an independent external cohort	Model-parameter fitting rather than probabilistic calibration	Mechanistic physiological parameters provide interpretability	Limited to national-level cyclists	Predictive individualised model	Level 3: externally anchored/field-validated predictive twin
11	Boillet (2025)	Modelling study	Rowing-performance simulations for 2,000 m and proposed 1,500 m Olympic distances	Simulation study; athlete/model parameter source NR in accessible abstract	Rowing	Energetic and physiological performance parameters	Physiology-based energetic simulation/digital-twin modelling	Comparative simulation and sensitivity analysis	Energy-system contribution and sensitivity to physiological parameters	Reducing race distance increased the relative non-oxidative energy contribution by about 5%, while	No independent prospective validation reported	Not applicable	Mechanistic model is interpretable	Not assessed	Simulation for competition-planning questions	Level 1–2: simulation/internal validation

										sensitivity to peak oxygen uptake remained comparable .						
12	Naughton (2024)	Perspective/review	Sports-science and sports-medicine multidisciplinary teams	Opinion article; no empirical sample	Sport	Human-factors and sociotechnical implementation evidence	Generative AI, automation and human-AI teaming	Expert commentary using historical and contemporary examples	Implementation opportunities, risks and work-system implications	AI may augment or replace selected team functions, but successful adoption requires sociotechnical design, human oversight and explicit management of unintended consequences.	Not applicable	Not applicable	Transparency and role clarity emphasised	Workforce and stakeholder effects discussed	Conceptual implementation guidance	Level 0: conceptual/governance evidence
13	Dergaa (2024)	Expert evaluation study	GPT-4 exercise-prescription outputs evaluated against professional expectations	Prompt/output evaluation ; exact number NR in accessible abstract	Exercise prescription	Text-based exercise prescriptions	OpenAI GPT-4	Critical expert evaluation of generated exercise recommendations	Appropriateness, safety and quality of exercise prescription	GPT-4 generated plausible exercise advice but required professional verification because specificity, safety and contextualisation were inconsistent.	No independent model/site validation	Not applicable	Generated rationale available but not mechanistically explainable	Limited contextual and subgroup personalisation	Expert-evaluated generative-AI output	Level 1–2: feasibility/internal expert evaluation

14	Wang (2024)	Review	Published literature on AI applications and teacher preparation in physical education	NR (comprehensive review)	Physical education	Education al, assessment and personalised-learning data	AI-supported instruction, analytics and adaptive systems	Comprehensive narrative review	Teaching applications , benefits, risks and training needs	AI can support personalised PE and data-informed instruction, but teacher competence, ethics, privacy and equitable access are prerequisites.	Not applicable to review	Not applicable	Transparency discussed as a governance requirement	Digital equity highlighted	Evidence-synthesis and policy level	Contextual evidence synthesis
15	Zhong (2025)	Systematic review	85 empirical studies published from January 2014 to November 2024	85 studies	Physical education	Instructional-design, visualisation, assessment, feedback and health-monitoring data	Adaptive platforms, virtual simulation, multimodal assessment and digital-intelligent systems	PRISMA systematic review	Application scenarios, educational outcomes and implementation barriers	Digital-intelligent technologies form a closed instructional loop but face algorithmic, privacy and digital-equity barriers.	Not applicable to review	Not applicable	System transparency and governance identified as needs	Unequal access and educational-stage disparities identified	Evidence synthesis across educational stages	Contextual evidence synthesis
16	Makaruk (2025)	Validation study	236 participants: 126 primary-school students aged 7–14 and 110 university sport students aged 20–21	236 participants	Motor skills / jumping rope	Mobile-app video of jumping-rope performance	AI model scoring five process-oriented criteria in the FUS test	Concurrent validity, inter-rater and intra-rater reliability against expert ratings; 3-week reassessment	ICC 0.96; weighted kappa 0.87; criterion kappas 0.83–0.87; AI intra-rater ICC/kappa 1.00	AI scoring showed excellent agreement with experts and improved further after expert correction.	No independent external-site validation	Not applicable	Five explicit movement criteria provide interpretable scoring	Included school and university groups of both sexes; single national setting	Validated educational assessment application	Level 3: independent expert-reference validation
17	He (2024)	Small intervention study	10 university students used for	10 students	400-m running	Visual-analysis and interactive	Interactive AI based	Pre-post teaching application with	400 m running performance and 10-	Interactive visual-AI teaching was	No independent cohort	Not reported	Not reported	Small single-institution	Small educational field	Level 4: small field evaluation without

			pre-post 400 m running and interactive-system evaluation		instruction	e teaching data	on visual analysis	performance comparison and user-satisfaction assessment	point user-satisfaction indicators	associated with improved 400 m performance and high user satisfaction ; 70% reported better understanding of physical-education content.				student sample	application	demonstrated long-term outcome impact
18	Ma (2025)	Empirical predictive study	800 Chinese university football players from a public Kaggle dataset	800 players; 18 predictors	University football	Basic information, training, physical-fitness and lifestyle variables	Ten ML algorithms; best model support vector machine; SHAP interpretation	Single-dataset train/test modelling	Accuracy 95.6%; F1 95.7%; ROC-AUC 99.2%	Stress, sleep duration and balance were leading predictors, but the single public dataset and absence of external validation substantially limit generalisability.	No	Not independently assessed; probability comparison reported	Yes—SHAP	Single-country university dataset	Retrospective technical validation	Level 2: retrospective internal validation
19	An (2024)	Scoping review	Published studies applying AI to physical-activity interventions and research	NR in accessible abstract	Physical activity	Wearable, app, behavioural and intervention data	Machine learning, deep learning and reinforcement learning	Scoping review of PubMed, Web of Science, Cochrane Library and EBSCO	Application type, algorithm and physical-activity outcome	AI was used for promotion, prediction and pattern recognition ; several AI-driven	Not applicable to review	Not systematically reported	Limited	Population heterogeneity noted	Evidence-synthesis level	Contextual evidence synthesis

										interventions improved physical-activity outcomes, but methodological standardisation was limited.						
20	Dolson (2022)	Systematic review	Published studies using wearable sensors to estimate core body temperature	NR in accessible abstract	Exercise / heat illness	Wearable physiological and environmental sensor data	Regression and ML-based core-temperature estimation models	Systematic review	Prediction error, agreement and validation method	Wearable estimation is promising, but sensor combinations, reference standards and validation procedures were heterogeneous.	Limited across reviewed studies	Variable/not consistently reported	Limited	Participant and environmental representation heterogeneous	Evidence synthesis of laboratory and field models	Contextual evidence synthesis
21	Kautz (2018)	Systematic review	52 studies of sport-specific movement recognition	52 studies	Sport-specific movement	IMU and computer-vision data	Predominantly supervised classification; kernel SVM in 53% of IMU and 50% of vision studies; 12 CNN studies and one LSTM application	Systematic review of model development and evaluation	Classification performance and methodological characteristics	High reported recognition performance was accompanied by major variation in preprocessing, experimental design and evaluation methods.	Limited and inconsistent	Not generally reported	Limited	Sport and participant representation heterogeneous	Evidence synthesis of technical models	Contextual evidence synthesis

22	Mundt (2022)	Empirical modelling study	14 participants completing running, sidestepping and cross-over tasks	14 participants; 1,474 usable movement samples	Human movement / sports biomechanics	2D keypoints from AlphaPose, BlazePose and OpenPose; force-platform GRF reference	Artificial neural networks for 3D ground-reaction-force estimation	Three test cases assessing model accuracy and cross-pose-estimator interchangeability against force-plate ground-reaction forces	Keypoint detection rate; GRF waveform, loading-rate and peak-force estimation	AlphaPose and OpenPose keypoints were partly interchangeable for GRF estimation, whereas BlazePose had lower detection and interchangeability.	No independent cohort	Not reported	Input keypoints and biomechanical outputs are inspectable	Participant composition and broader demographic representativeness were limited	Laboratory technical validation	Level 2: retrospective/laboratory validation
23	Johnson (2019)	Empirical cohort/predictive study	30 participants completing sports-related movements in laboratory motion-capture and force-plate testing	30 participants	Team-sport athletes	Wearable-sensor workload and movement data	Machine/deep-learning injury-risk monitoring	Pre-trained CaffeNet CNN with multivariate regression from marker-based motion capture to 3D knee-joint moments across three sports-related movements	Correlation between CNN-estimated and reference-model knee-joint moments; strongest mean correlation 0.8895	Deep learning showed feasibility for portable estimation of knee-joint loading; the strongest overall mean correlation was 0.8895 during the initial 33% of stance in sidestepping.	No independent external cohort; laboratory criterion comparison	Not reported	Limited	Cohort representation NR	Laboratory technical validation directed toward on-field use	Level 2–3: criterion-validated technical model
24	Ishida <i>et al.</i> (2024)	Empirical validation study	18 healthy male participants; age 23.0 ± 1.6 years	18 participants	Sports physical therapy; single-leg landing	Two-dimensional video, pose-estimation AI, 3D optical motion	Pose-estimation AI applied to 2D video for centre-of-mass tracking	Concurrent comparison of 2D-AI and 3D motion-capture estimates against force-plate	Peak vertical ground reaction force and agreement/error versus	Pose-estimation AI using ordinary 2D video provided a feasible estimate of	No independent external cohort; concurrent criterion validation	Not applicable to continuous biomechanical	Biomechanical centre-of-mass and force-estimation	Healthy young men only; no female, injured or elite-	Controlled laboratory validation with clinically accessible	Level 2–3: laboratory criterion validation

					biomechanics	capture and force-plate data	and VGRF estimation	VGRF during single-leg landing	force-plate reference	vertical ground reaction force during single-leg landing, although reference-based validation and task-specific constraints remain important.	in 18 participants	estimation	pipeline is directly interpretable	athlete validation	video input	
25	Lahkar et al. (2022)	Empirical validation study	Three elite boxers performing standardised shadow-boxing sequences	3 elite boxers	Elite boxing; markerless motion analysis	Synchronous 10-camera video and 12-camera optoelectronic marker-based motion capture	Commercial markerless computer-vision motion-capture system	Concurrent comparison with a marker-based optoelectronic reference system	Three-dimensional shoulder, elbow and wrist joint-centre position differences	Markerless estimates were closer for shoulder and wrist joint centres than for the elbow; elbow errors exceeded 3 cm, whereas shoulder and wrist errors were generally below 2.5 cm.	No independent cohort; direct criterion validation against marker-based motion capture	Not applicable to coordinate estimation validation	Joint-centre errors and sport-specific movements are directly interpretable	Only three elite boxers; participant sex and broader athlete diversity were limited	Controlled sport-specific laboratory validation	Level 2–3: laboratory criterion validation
26	Adlou (2025)	Systematic review	Published studies of motion-capture technologies for athletic	NR in accessible abstract	Athletic performance	Marker-based, markerless, inertial and video motion-	Computer vision, pose estimation and analytical pipelines	Systematic review	Validity, reliability, applications and practical constraints	Motion capture supports performance analysis across sports,	Not applicable to review	Not applicable	Varies by technology	Population reporting heterogeneous	Evidence-synthesis level	Contextual evidence synthesis

			performanc e			capture data				while cost, setup, ecological validity and cross- system comparabili ty remain important constraints.						
27	Schortge n (2025)	Empirical computer- vision study	22 elite amateur boxing fighters recorded during 18 real- competition rounds	22 fighters; 18 rounds	Boxing	Single- camera monocula r boxing video	Self- supervised detection and markerless 2D position tracking	Real- competition robustness testing of monocular markerless tracking under occlusion	Position/traj ectory tracking performanc e	The method enabled planar boxer- position tracking from monocular video despite frequent occlusion, but evidence remains a technical validation in a specialised setting.	No independen t external dataset identified	Not applica ble	Detectio n/tracki ng pipeline describe d	Elite amateur boxing only	Authentic competiti on-field evaluatio n	Level 4: field evaluation without demonstr ated decision or athlete- outcome impact
28	Kim (2025)	Systemati c review	25 studies selected from 397 records	25 studies	Sport governan ce	Ethical, governan ce and sport- technolog y evidence	AI applicatio ns across sport	Systematic scoping review of nine databases; preregistered protocol	Four ethical domains and frequency of concerns	Privacy/dat a ethics appeared in 22 studies, fairness/bia s in 15, transparen cy/explaina bility in 13 and accountabil ity in 8.	Not applicable	Not applica ble	A primary ethical domain	Fairness and underrepr esentatio n were central findings	Governan ce evidence synthesis	Contextua l governanc e evidence

29	Musat (2024)	Review	Published sports-injury prediction studies across multiple sports	NR in accessible abstract	Multiple sports	Training load, biomechanical, physiological and athlete-profile data	Random forests, CNNs, ANNs and other ML/DL methods	Comprehensive review	Predictive performance, model type and application area	AI can support injury-risk estimation, but inconsistent datasets, validation, interpretability and prospective testing restrict clinical translation.	Not applicable to review; limited in reviewed evidence	Infrequently reported	Often limited	Dataset diversity identified as a concern	Evidence synthesis	Contextual evidence synthesis
30	Vitale (2025)	Expert evaluation study	Athlete sleep and jet-lag questions evaluated by domain experts across large language models	Exact prompt/model count NR in accessible abstract	Athletes / travel recovery	Text responses on sleep and jet-lag advice	Multiple large language models	Blinded/expert assessment of answer quality and trustworthiness	Accuracy, appropriateness, safety and expert ratings	LLMs produced useful but inconsistent advice; expert oversight remained necessary before athlete-facing use.	Model comparison rather than independent clinical validation	Not applicable	Natural-language rationale available, but source attribution/reliability variable	Subgroup personalisation limited	Expert-evaluated generative-AI application	Level 2: controlled expert evaluation
31	Gao (2025)	Narrative/conceptual review	Perspective article; no empirical participant sample	Not applicable — perspective article	Higher education	Teaching, evaluation and personalised-training data	AI-supported teaching models and evaluation systems	Perspective synthesis of AI-enabled teaching, evaluation and personalised-training models	Not applicable—no original empirical outcome	The article argues that AI can support personalised PE and public-health education, but independent validation and detailed methodological	Not applicable	Not applicable	Not reported	Equity considerations not fully evaluated	Conceptual and policy-oriented evidence	Level 0: conceptual/perspective evidence

										reporting are limited.						
32	Raju (2026)	Empirical predictive study	Multi-sport university athletes from a publicly available athlete injury and performance dataset	Public dataset; exact observation count NR in accessible abstract	Athletes	Workload, recovery, demographic, physiological and anthropometric variables	Supervised ML; random forest best-performing; SHAP-based interpretation	Train/test modelling on a single public dataset	Classification performance and SHAP feature importance	Workload, recovery and athlete characteristics supported injury prediction, but single-dataset design requires external and prospective validation.	No	Predicted-versus-observed probabilities presented; independent calibration not established	Yes—SHAP	University multi-sport dataset; broader representation not established	Retrospective technical validation	Level 2: retrospective internal validation

Table S3. SPORT-AI operational coding manual

Dimension	Operational definition
Scientific validity	Appropriateness of design, outcome definition, data quality, and analytical procedure
Predictive performance	Reported discrimination, classification, agreement, or estimation accuracy
Out-of-sample validation	Temporal, organisational, geographic, device, or independently assembled dataset testing; excludes within-dataset cross-validation
Real-world utility	Evaluation in authentic sporting, clinical, educational, or organisational workflows
Transparency	Model, preprocessing, prompts, version, code/data availability, explainability, and reproducibility
Athlete rights	Consent, privacy, fairness, autonomy, ownership, contestability, and human oversight
Implementation readiness	Workflow fit, interoperability, stakeholder acceptance, maintenance, governance, and deployment evidence

Table S4. Study level SPORT-AI translational assessment

ID	Study	SPORT-AI source coding	Scientific validity	Predictive performance	Out-of-sample validation	Real-world utility	Transparency	Athlete rights	Implementation readiness
1	Claudino (2019)	Source-verified coding	Contextual/design-dependent	Not applicable	Not applicable	Evidence synthesis; predominantly retrospective technical studies	Model interpretability not systematically established	Pooled evidence was strongly male dominated (97%)	Low/emerging
2	Reis (2024)	Source-verified coding	Contextual/design-dependent	Not applicable	Not applicable	Evidence-synthesis level	Explainability identified as a central challenge	Bias and representativeness discussed conceptually	Low/emerging
3	Souaifi (2025)	Source-verified coding	Contextual/design-dependent	Not applicable	Not applicable	Evidence synthesis with predominantly technical validation	Interpretability identified as a persistent limitation	Population representation inconsistently reported	Low/emerging
4	Van Eetvelde (2021)	Source-verified coding	Contextual/design-dependent	Not applicable	Not applicable	Retrospective predictive evidence	Variable and usually limited	Small and heterogeneous athlete samples	Low/emerging
5	Jauhiainen (2022)	Source-verified coding	High/adequate for design	Reported	Criterion or out-of-sample status reported separately	Retrospective internal validation	Risk-factor contribution examined, but no full clinical explanation framework	Female elite handball and soccer athletes only	Low/emerging
6	Seshadri (2021)	Source-verified coding	Contextual/design-dependent	Not applicable	Not applicable	Evidence-synthesis level	Human interpretation remains necessary	Representation not systematically appraised	Low/emerging
7	Chidambaram (2022)	Source-verified coding	Contextual/design-dependent	Not applicable	Not applicable	Evidence synthesis; mostly technical/observational applications	Model transparency inconsistently reported	Sport and participant representation heterogeneous	Low/emerging
8	Gámez Díaz (2020)	Source-verified coding	Contextual/design-dependent	Not applicable	Not applicable	Conceptual and prototype systems	System transparency discussed conceptually	User inclusion not systematically quantified	Low/emerging

9	Fister Jr. (2021)	Verified clean-core coding	Contextual/design-dependent	Not applicable	Not applicable	Working prototype	Conceptual cognitive model described	Not assessed	Low/emerging
10	Boillet (2024)	Source-verified coding	High/adequate for design	Reported	Criterion or out-of-sample status reported separately	Predictive individualised model	Mechanistic physiological parameters provide interpretability	Limited to national-level cyclists	Moderate
11	Boillet (2025)	Source-verified coding	Contextual/design-dependent	Not applicable	Criterion or out-of-sample status reported separately	Simulation for competition-planning questions	Mechanistic model is interpretable	Not assessed	Low/emerging
12	Naughton (2024)	Source-verified coding	Contextual/design-dependent	Not applicable	Not applicable	Conceptual implementation guidance	Transparency and role clarity emphasised	Workforce and stakeholder effects discussed	Low/emerging
13	Dergaa (2024)	Source-verified coding	Contextual/design-dependent	Not applicable	Criterion or out-of-sample status reported separately	Expert-evaluated generative-AI output	Generated rationale available but not mechanistically explainable	Limited contextual and subgroup personalisation	Low/emerging
14	Wang (2024)	Source-verified coding	Contextual/design-dependent	Not applicable	Not applicable	Evidence-synthesis and policy level	Transparency discussed as a governance requirement	Digital equity highlighted	Low/emerging
15	Zhong (2025)	Source-verified coding	Contextual/design-dependent	Not applicable	Not applicable	Evidence synthesis across educational stages	System transparency and governance identified as needs	Unequal access and educational-stage disparities identified	Low/emerging
16	Makaruk (2025)	Source-verified coding	High/adequate for design	Reported	Criterion or out-of-sample status reported separately	Validated educational assessment application	Five explicit movement criteria provide interpretable scoring	Included school and university groups of both sexes; single national setting	Moderate

17	He (2024)	Source-verified coding	Contextual/design-dependent	Not applicable	Criterion or out-of-sample status reported separately	Small educational field application	Not reported	Small single-institution student sample	Moderate
18	Ma (2025)	Source-verified coding	High/adequate for design	Reported	No independent external validation	Retrospective technical validation	Yes—SHAP	Single-country university dataset	Low/emerging
19	An (2024)	Source-verified coding	Contextual/design-dependent	Not applicable	Not applicable	Evidence-synthesis level	Limited	Population heterogeneity noted	Low/emerging
20	Dolson (2022)	Source-verified coding	Contextual/design-dependent	Not applicable	Not applicable	Evidence synthesis of laboratory and field models	Limited	Participant and environmental representation heterogeneous	Low/emerging
21	Kautz (2018)	Source-verified coding	Contextual/design-dependent	Not applicable	Not applicable	Evidence synthesis of technical models	Limited	Sport and participant representation heterogeneous	Low/emerging
22	Mundt (2022)	Source-verified coding	High/adequate for design	Reported	Criterion or out-of-sample status reported separately	Laboratory technical validation	Input keypoints and biomechanical outputs are inspectable	Participant composition and broader demographic representativeness were limited	Low/emerging
23	Johnson (2019)	Source-verified coding	High/adequate for design	Reported	Criterion or out-of-sample status reported separately	Laboratory technical validation directed toward on-field use	Limited	Cohort representation NR	Low/emerging
24	Ishida (2024)	Source-verified coding	High/adequate for design	Reported	Criterion or out-of-sample status reported separately	Controlled laboratory validation with clinically accessible video input	Biomechanical centre-of-mass and force-estimation pipeline is directly interpretable	Healthy young men only; no female, injured or elite-athlete validation	Low/emerging

25	Lahkar (2022)	Source-verified coding	High/adequate for design	Reported	Criterion or out-of-sample status reported separately	Controlled sport-specific laboratory validation	Joint-centre errors and sport-specific movements are directly interpretable	Only three elite boxers; participant sex and broader athlete diversity were limited	Low/emerging
26	Adlou (2025)	Source-verified coding	Contextual/design-dependent	Not applicable	Not applicable	Evidence-synthesis level	Varies by technology	Population reporting heterogeneous	Low/emerging
27	Schortgen (2025)	Source-verified coding	Contextual/design-dependent	Reported	Criterion or out-of-sample status reported separately	Authentic competition-field evaluation	Detection/tracking pipeline described	Elite amateur boxing only	Moderate
28	Kim (2025)	Source-verified coding	Contextual/design-dependent	Not applicable	Not applicable	Governance evidence synthesis	A primary ethical domain	Fairness and underrepresentation were central findings	Low/emerging
29	Musat (2024)	Source-verified coding	Contextual/design-dependent	Not applicable	Not applicable	Evidence synthesis	Often limited	Dataset diversity identified as a concern	Low/emerging
30	Vitale (2025)	Source-verified coding	Contextual/design-dependent	Not applicable	Criterion or out-of-sample status reported separately	Expert-evaluated generative-AI application	Natural-language rationale available, but source attribution/reliability variable	Subgroup personalisation limited	Low/emerging
31	Gao (2025)	Source-verified coding	Contextual/design-dependent	Not applicable	Not applicable	Conceptual and policy-oriented evidence	Not reported	Equity considerations not fully evaluated	Low/emerging
32	Raju (2026)	Source-verified coding	High/adequate for design	Reported	No independent external validation	Retrospective technical validation	Yes—SHAP	University multi-sport dataset; broader representation not established	Low/emerging