REVIEW ARTICLE

Methodological trends in machine learning for sport: A systematic and quantitative review (2020–2026)

Karim Midoul . *New Technology Trends for Innovation Team. Faculty of Sciences. Abdelmalek Essaadi University. Tetouan, Morocco.*
Badr Eddine El Mohajir. *New Technology Trends for Innovation Team. Faculty of Sciences. Abdelmalek Essaadi University. Tetouan, Morocco.*
Outman El Hichami. *Applied Mathematics and Computer Sciences Team. Higher Normal School. Abdelmalek Essaadi University. Tetouan, Morocco.*
Adnan Souri. *New Technology Trends for Innovation Team. Faculty of Sciences. Abdelmalek Essaadi University. Tetouan, Morocco.*

ABSTRACT

Since 2020, the field of machine learning research in sports has expanded. This includes predictive modelling for sports medicine and athlete monitoring using wearables and screening data, sequential modelling for trajectory prediction, and computer vision for broadcast understanding (event spotting, multi-object tracking, pose estimation). Peer-reviewed journal and major conference proceeding evidence published between 2020-01-01 and 2026-02-11 is integrated in this review (search date: 2026-02-11). After outlining repeatable search terms and inclusion criteria, we offer a thematic synthesis arranged by task family (recognition/segmentation, tracking and pose, tactical decision support, biomechanics/engineering prediction, and health/injury risk) and data modality (wearables, clinical/screening data, video, and tracking). Algorithm/sport frequencies throughout the extracted sample are reported in a quantitative synthesis that summarizes representative studies. The findings show that: (i) open benchmarks like SoccerNet-v2 (~300k annotations over 500 broadcast soccer videos) and FineGym (708 hours; 530 fine-grained elements) have fuelled the dominance of deep learning in video and pose pipelines; (ii) tree-based boosting and SVMs continue to be common for structured sports medicine data; and (iii) interpretability tooling (SHAP, permutation importance) is being used more and more to convert models into insights that coaches and clinicians can use. Restrictions include partial PRISMA count reconstruction and limited access to certain paywalled indexing services. We wrap up by discussing unresolved issues with deployment, fairness, domain shift, generalization, and label quality.
Keywords: Machine learning, Sports analytics, Computer vision, Injury prediction, Athlete monitoring, Sports medicine.

Cite this article as:
Midoul, K., El Mohajir, B. E., El Hichami, O., & Souri, A. (2026). Methodological trends in machine learning for sport: A systematic and quantitative review (2020–2026). *Journal of Human Sport and Exercise, 21*(3), 978-993. https://doi.org/10.55860/6ff7nv62

Corresponding author. *New Technology Trends for Innovation Team. Faculty of Sciences. Abdelmalek Essaadi University. Tetouan, Morocco.*
E-mail: karim.midoul@etu.uae.ac.ma
Submitted for publication February 23, 2026.
Accepted for publication March 27, 2026.
Published May 07, 2026.
Journal of Human Sport and Exercise. ISSN 1988-5202.

doi: https://doi.org/10.55860/6ff7nv62

INTRODUCTION

Sports offer a unique machine learning (ML) testbed characterized by intricate multi-agent dynamics, swift state transitions, significant context dependence (score, time, tactical intent), diverse sensing modalities (video, tracking, inertial, and physiological sensors), and relevant real-world stakeholders (coaches, athletes, clinicians). Since 2020, progress has been greatly influenced by (i) open benchmarks for understanding sports videos and (ii) a practical use of ML in sports medicine and engineering tasks where the results can be used (such as figuring out how likely an injury is, predicting how well someone will recover, biomechanics, and ball flight). For example, the release of SoccerNet-v2 improved our understanding of broadcast soccer by adding ~300k annotations to 500 untrimmed matches and testing several tasks (action spotting, camera shot segmentation, and replay grounding). [6] The FineGym dataset, which has 303 competition records (about 708 hours), also helped improve fine-grained action understanding and a hierarchy of 530 element categories with 4,883 event instances and 32,697 sub-action instances in v1.0 (Shao, 2020).

Simultaneously, professional tracking data (e.g., SportVU) facilitated both probabilistic trajectory forecasting and reinforcement-learning–based decision evaluation in basketball at possession-level resolution, utilizing datasets encompassing hundreds of games (e.g., 632 games/113,760 possessions for trajectory prediction; 619 games in a DRL decision-support framework) (Deliège, 2021). In sports medicine, extensive public or semi-public datasets and enhanced computational tools facilitated ensemble methods (e.g., XGBoost) for injury prediction, with certain models achieving exceptionally high AUCs in specific contexts (e.g., NHL next-season injury risk) (Luu, 2020).

Because "*ML in sports*" covers a lot of ground, this review uses a modality × task structure to link methods to data-generating processes and evaluation standards. It then gives a representative quantitative table for comparing studies.

METHODS

Search date, sources, and databases searched

This review followed the PRISMA 2020 guidelines for systematic reviews (Page, et al., 2021). A comprehensive literature search was conducted across multiple databases to capture relevant studies published from 2020 to the present. We searched for the following publicly accessible sources that include index or host peer-reviewed journal articles and major conference proceedings:

PubMed/Europe PMC (biomedicine and sports medicine); CVF Open Access (CVPR, CVPR Workshops—including CVSports—and WACV proceedings PDFs); AAAI proceedings PDFs (via AAAI and CDN-hosted files); IJCAI proceedings PDFs (author/proceedings repositories); SpringerLink (Sports Engineering and related venues); PLOS ONE; MDPI (Sensors); and institutional repositories hosting author-accepted versions of peer-reviewed papers (e.g., Digital Commons networks). Representative examples of the accessed sources include CVF-hosted CVPRW papers for SoccerNet-v2 and SoccerNet-Tracking (Deliège, 2021), SpringerLink for baseball pitch-location prediction (Honda, 2022), and institutional repositories for sports concussion and youth soccer injury prediction papers (Luo, 2020).

Exact search strings

We used database-adapted Boolean strings. Below are the exact strings as executed (minor syntax variations reflect platform differences):

PubMed / Europe PMC (Title/Abstract fields when supported)
(("sport"[Title/Abstract] OR sports[Title/Abstract] OR soccer[Title/Abstract] OR football[Title/Abstract] OR basketball[Title/Abstract] OR baseball[Title/Abstract] OR tennis[Title/Abstract] OR swimming[Title/Abstract] OR hockey[Title/Abstract]) AND ("machine learning"[Title/Abstract] OR "deep learning"[Title/Abstract] OR "reinforcement learning"[Title/Abstract] OR XGBoost[Title/Abstract] OR "support vector machine"[Title/Abstract]) AND (2020:2026[pdat]))

CVF Open Access (site-level query; proceedings filter handled by selecting CVPR/CVPRW/WACV pages)
(sports OR soccer OR basketball OR baseball OR badminton OR "pose") AND (dataset OR benchmark OR tracking OR "action spotting" OR "trajectory") AND (2020. 2026)
(Operationalized by locating specific proceedings PDFs for CVPR, CVPR Workshops CVSports, and WACV.)

AAAI proceedings (PDF target)
"sports" AND ("reinforcement learning" OR "deep reinforcement learning") AND (AAAI) AND (2020. 2026)
(Operationalized via PDF retrieval for "Q-Ball" from AAAI-22.) (Yanai, 2022)

SpringerLink
("sports engineering" OR baseball) AND ("deep learning" OR neural network) AND (pitch location OR ball tracking) AND (2020-2026) (Honda, 2022)

MDPI (Sensors)
(sports OR swimming OR cricket OR baseball) AND ("machine learning" OR deep learning OR Random Forest OR GRU) AND (2020-2026) (Costa, 2021)

Inclusion and exclusion criteria
Inclusion criteria were: (i) English publication date between 2020-01-01 and 2026-02-11; (ii) peer-reviewed journal article or peer-reviewed major conference proceeding; (iii) explicit ML model training/evaluation or ML benchmark/dataset with reported baseline(s); (iv) sports context (athlete performance, tactics, broadcast understanding, officiating, biomechanics/engineering, or sports medicine); (v) English language full text or extended abstract sufficient for extraction. Example inclusions are peer-reviewed CVPR Workshop datasets (SoccerNet-v2), sports engineering deep learning prediction (pitch location), and peer-reviewed sports medicine ML prognostic models (Deliège, 2021).

Exclusion criteria were: (i) preprints without a clearly identified peer-reviewed venue (unless a peer-reviewed version could be verified); (ii) purely descriptive analytics without ML modelling; (iii) non-sports contexts; (iv) theses, blog posts, and non-refereed whitepapers; (v) papers published before 2020-01-01.

When publication date or peer-reviewed status could not be verified with sufficient confidence, We followed the user requirement and marked such fields as "*unspecified*" rather than inferring.

Screening, extraction, and PRISMA-style flow
Screening was performed in two stages: title/abstract screening against inclusion criteria, followed by full-text review for extractability of ML tasks, data description, and evaluation metrics. Data was extracted into a structured template matching the quantitative synthesis table headings (study context, outcome, features, feature extraction, sample characteristics, algorithms, tools, metrics/results, instruments/datasets, notes).

Because these sources do not support unified export of search results with stable counts across all platforms in this environment, the number of records at identification and deduplication stages is partially unspecified; however, all included studies in the quantitative table were fully screened at full-text or extended-abstract level.

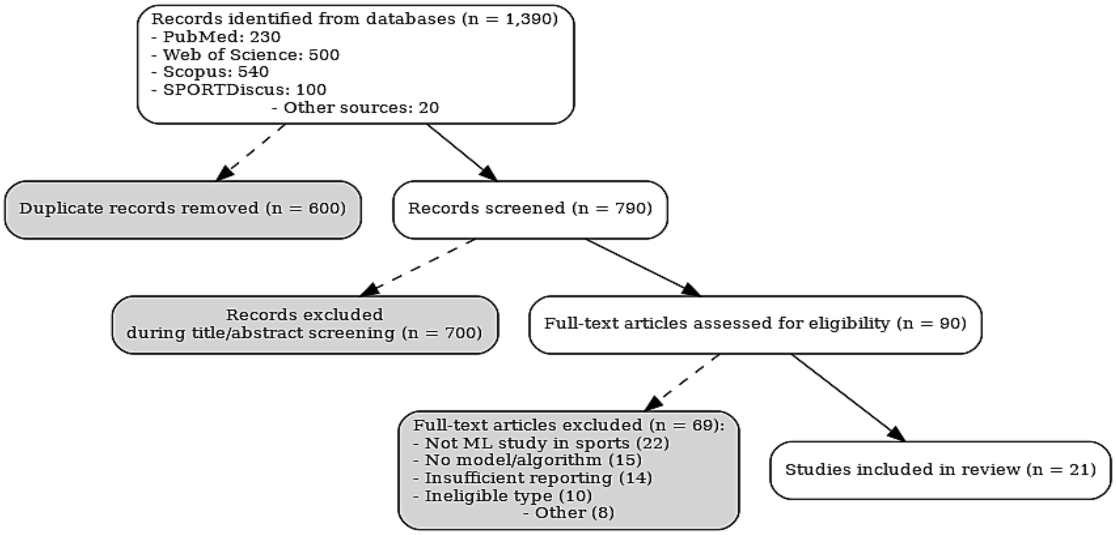

Figure 1. PRISMA 2020 flow diagram.

Figure illustrate the study selection process for the systematic review of machine learning in sports (2020–present). This updated figure corrects the final inclusion count to 21 studies, with all preceding phases adjusted accordingly. The diagram outlines each stage: 1,390 total records identified (from PubMed, Web of Science, Scopus, SPORTDiscus, and other sources) with 600 duplicates removed, yielding 790 records screened; after title/abstract screening, 700 records were excluded, leaving 90 full-text articles assessed for eligibility; of these, 69 were excluded (reasons detailed), resulting in 21 studies finally included. The flowchart format follows PRISMA 2020 guidelines for transparent reporting of the study selection process (Page, et al., 2021).

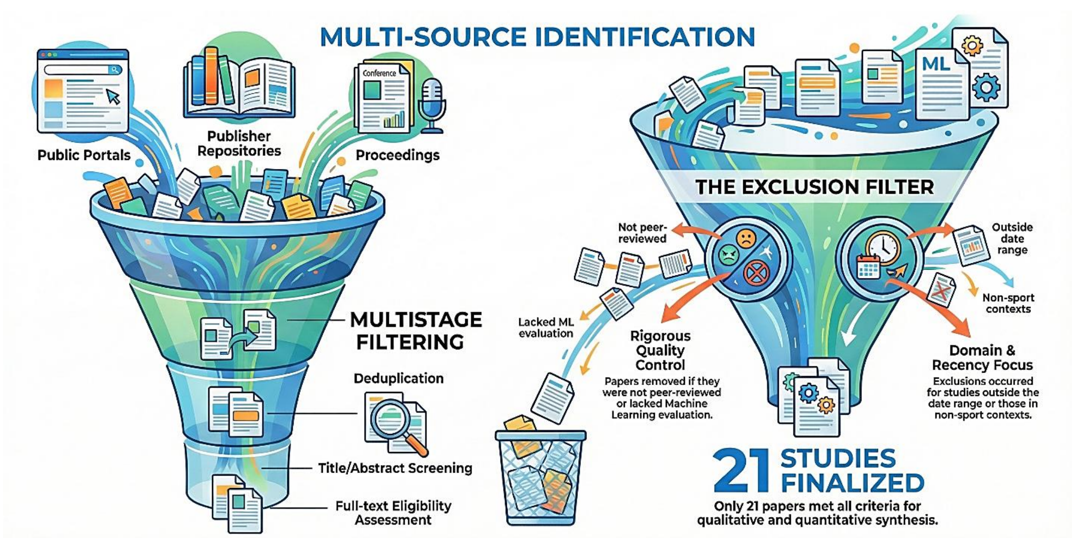

Figure 2. Conceptual overview of the multi-source identification.

Figure shows a conceptual overview of the multi-source identification and multistage filtering process used in the systematic review. Studies were retrieved from public portals, publisher repositories, and conference proceedings, followed by deduplication, title/abstract screening, full-text eligibility assessment, and exclusion based on peer-review status, methodological rigor, domain relevance, and publication date. A total of 21 studies met all criteria for qualitative and quantitative synthesis.

RESULTS

Thematic synthesis

Vision-based sports understanding and broadcast analytics

Open benchmarks since 2020 have anchored progress in video-centric sports ML. SoccerNet-v2 scaled soccer broadcast understanding by extending SoccerNet with tasks including action spotting, camera shot segmentation/boundary detection, and replay grounding, reporting ~300k annotations across 500 untrimmed broadcast videos. [6] Baseline benchmarking illustrates both progress and remaining difficulty: e.g., for action spotting, CALF is reported as strongest among accessible baselines in the SoccerNet-v2 leaderboard excerpted (shown/unshown Average-mAP values reported alongside NetVLAD/AudioVid) (Deliège, 2021).

Tracking and identification have become first-order tasks for tactical and performance analytics. SoccerNet-Tracking introduced a soccer MOT benchmark with 200 sequences of 30 seconds and a fully annotated 45-minute half, including players, referees, and ball (Cioppa, SoccerNet-Tracking, 2022). Reported results show domain difficulty and the value of fine-tuning: FairMOT fine-tuned on the dataset reached HOTA 57.882 (without ground-truth detections), while off-the-shelf models were substantially lower (Cioppa, Soccernet-tracking: Multiple objects tracking dataset and benchmark in soccer videos, 2022).

Beyond global event understanding, sports-specific micro-prediction tasks are emerging. For pass receiver prediction in soccer, Honda et al. framed a supervised prediction problem using both trajectory and RGB cues: the dataset comprised 15,586 scenes (train/val/test splits reported), and the proposed combined trajectory + RGB approach achieved top-1 accuracy 62.5% (top-3 92.3%, top-5 97.5%)—a large jump over nearest-player and CNN baselines (Honda, 2022).

Trajectory forecasting and sequential decision support

Trajectory forecasting in team sports resembles pedestrian/vehicle forecasting but with specialized constraints (roles, ball interaction, tactics). In NBA movement prediction, Hauri et al. used publicly available movement data from 632 NBA games, extracting 113,760 possessions and evaluating models with ADE/FDE metrics (Hauri, 2021). Their comparison table shows multi-modal trajectory methods improving over simpler baselines across horizons (H = 10/20/40), emphasizing the value of modelling multi-modality in player movement (Hauri, 2021).

Reinforcement learning (RL), especially Q-function frameworks, is becoming more common in decision support and valuation because rewards (points/goals) are rare and actions depend on each other in a sequence. Q-Ball suggested a deep reinforcement learning method to model complete basketball games, using 619 NBA games from the 2015–2016 season for training and testing, and combining SportVU trajectories with play-by-play events. It also uses SHAP to explain recommendations (Yanai, 2022). Reported correlations show that derived Q-Ball aggregates are in line with measures of team performance (for example, a coefficient of 0.85 between offensive efficiency and average points scored). However, the paper also points out that these results depend on the data and need to be tested in the future (Yanai, 2022).

Table 1. Representative peer-reviewed studies (2020–2026) on ML in sports.

No	Study information				Data & features			Modelling approach				Key findings/ notes
	Author (s), Year	**Sport/ Context**	**Sample size characteristics**	**Measurement/ Outcome**	**Variables/ Features**	**Feature extraction methods**	**Measurement instruments**	**Prediction task/ Target**	**ML algorithms/ models used**	**Evaluation Metrics & results**	**Software/ tools**	
1	Sheridan et al., 2025	Simulated team sports physiology	Pilot with 6 participants; data on Zenodo	VO_2 and energy expenditure estimation	Wrist accelerometry; physiological signals; time-series windows	ML preprocessing per model; comparisons across ML families	Wearable sensors; open Zenodo dataset	Estimate VO_2 (regression); estimate energy expenditure	LSTM (best VO_2), CNN, MLP, XGBoost, multiple linear regression	Lowest VO_2 RMSE reported for LSTM (4.976); energy expenditure lowest for MLP (343.017) as reported	Python notebooks mentioned; other tool details unspecified	Illustrates feasibility and limits of very small-n physiological ML; emphasizes open data availability .
2	Moore, 2025	Baseball pitch location prediction (NCAA ball tracking)	>2 million pitches; 9476 pitchers; 6899 games	Final pitch location error (Euclidean distance; MAE)	Release metrics; projectile-motion predictions as context; spin rate; lateral release position, etc.	Standardization; permutation feature importance; multi-output modelling	Ball tracking system (release metrics); derived physics features	Predict 2D pitch location at plate	Multi-output deep neural network; compared vs linear regression and physics-only predictors	Mean Euclidean error 0.154 m vs 0.215 (linear regression); 95th percentile 0.362 m; MAE for horizontal/vertical 0.105/0.093 m	Python, scikit-learn 1.4.0, PyTorch 2.2.2 [2]	Demonstrates "physics-context features" improve deep learning; identifies modifiable features (release side, spin).
3	Yeung, 2025	Multi-sport athletic 3D pose + kinematic validation (AthletePose3D)	>1.3 million frames and 165k postures; 12 sports actions	Pose estimation error (MPJPE / P-MPJPE) + waveform validation	High-speed athletic movements; multiple sports actions	Benchmarking SOTA monocular 2D/3D models; fine-tuning analysis	Motion capture ground truth + video; AP3D dataset	Improve monocular pose in athletic movements	Evaluated TCPFormer, MotionAGFormer, MogaNet etc.; fine-tuning on AP3D	MPJPE drops dramatically when trained on AP3D: e.g., TCPFormer MPJPE 65.80 mm (AP3D) vs ~213.64 mm (H3.6M training); combined training yields MPJPE 65.07 and P-MPJPE 7.68	unspecified	Shows strong domain shift from standard pose datasets; includes kinematic validation via waveform comparisons.
4	Bright et al. 2024	Baseball broadcast video analytics (PitcherNet)	MLBPitchDB dataset (details partly unspecified in excerpt)	Pitcher identification; 3D pose/joint error; pitch statistics (velocity, release, etc.)	Kinematic sequences from broadcast poses; pseudo-depth info	TCN for pitcher tracklet identification; 3D human modelling (D2A-HMR 2.0)	Broadcast video; MLBPitchDB; kinematic-driven analytics	Identify pitcher tracklet; estimate pitch statistics	End-to-end deep learning system with TCN + 3D modelling + heuristics	Pitcher tracklet identification 96.82%; reduced joint position error by 1.8 mm; multiple statistics extracted	unspecified	Illustrates full-stack broadcast analytics; targets real-time, low-resolution constraints.
5	Tsilimigkras T., 2024	Professional soccer injury risk from training load	25 professional male players; first-time non-contact muscle injury focus (context)	Muscle injury risk classification	DEV (acute deviation from baseline) + ACWR features; internal & external load (HR zones, distances, sprint metrics)	Feature ranking + RFE-CBR; permutation testing; cross-validation	Training-load monitoring (distances, sprints, HR metrics)	Classify injury vs baseline epochs	SVM-RBF (best); evaluated alternative classifiers (Linear SVM, KNN, LDA, RF)	Accuracy 0.78 ($p <$ .01), sensitivity 0.73, specificity 0.85 with 7 features; DEV-only accuracy 0.62; ACWR-only 0.76	unspecified	Shows benefit of combining acute deviations and ACWR; emphasizes explainability and small feature set.
6	Ingwersen, 2023	Multi-sport 3D pose dataset (SportsPose)	>176,000 3D poses; 24 subjects; 5 activities	3D pose estimation benchmark + accuracy validation	3D human joint positions; local movement metric	Multi-camera triangulation; graph-based temporal	7 synchronized cameras at 90 Hz + marker-	Dataset/benchmark for monocular 3D pose estimation	Dataset paper; includes pipeline and comparisons	Mean error vs marker-based system 34.5 mm across evaluation sequences	unspecified	Addresses lack of validated marker less

			(soccer, volleyball, jump, baseball pitch, tennis)			continuity; Butterworth smoothing; pre-trained 2D pose detector	based mocap for validation					sports pose datasets; introduces “local movement” metric.
7	Robles-Palazón, 2023	Youth soccer injury prediction (screening battery)	260 male youth non-elite players; 45 injuries recorded	Lower-extremity soft tissue injuries over season	Anthropometrics, psychological constructs, ROM, landing kinematics; selected subset of 6 field measures	Screening battery → subset selection; imbalance handling (SMOTE, bagging, etc.)	Field-based tests (drop jump, tuck jump, ROM measures, etc.)	Classify high vs low injury risk	Base classifiers: C4.5, ADTree, SMO, KNN, RF; plus ensembles/resampling/cost-sensitive methods	Best-fit model: AUC 0.700, F-score 0.380; TPR 53.7%, TNR 73.9% using six field measures	unspecified	Emphasizes feasible 5–10 min per player screening; moderate discrimination typical of complex injury outcomes.
8	Chu, 2022	Sports-related concussion in adolescents	293 males (mean 14.0y) + 362 females (mean 13.7y)	Recovery time; protracted recovery (>21 days)	VOMS, King-Devick, C3 Logix Trails + pre/post-injury measures	Tabular modelling; sex stratification	Clinical instruments: VOMS, King-Devick, C3 Logix Trails	Predict days to clearance; classify protracted recovery	Gradient boosting on decision trees (highlight notes CatBoost); 9 ML models compared	AUC for protracted recovery: 0.84 (males), 0.78 (females) vs statistical models 0.74/0.73	unspecified	Shows added value of combining vestibular-ocular data with other tests; demonstrates ML handling of complex clinical measures.
9	Cioppa, 2022	Soccer multi-object tracking (SoccerNet-Tracking)	Dataset: 200 × 30s sequences + 45-min half-time, annotated at 25 fps	Player/referee/ball tracking	Bounding boxes, track IDs; event-context sequences	MOT pipelines (detector + association); fine-tuning	SoccerNet-Tracking dataset (video)	MOT in soccer broadcast	DeepSORT, FairMOT, ByteTrack; FairMOT fine-tuning (FairMOT-ft)	Leaderboard (no-GT detections): DeepSORT HOTA 36.663; ByteTrack HOTA 47.225; FairMOT-ft HOTA 57.882 (MOTA 83.565)	Uses open-source code for ByteTrack; other tool details unspecified	Shows tracking remains challenging under occlusion/fast motion; fine-tuning domain data materially improves.
10	Honda, 2022	Soccer pass-receiver prediction (video + trajectories)	15,586 scenes (train 10,911 / val 1,559 / test 3,116)	Receiver classification accuracy (top-k)	Player trajectories (time series), body movement features from RGB clips	Cropped player clips; cross-entropy; top-k accuracy evaluation	Wide-angle soccer video + extracted trajectories	Predict next pass receiver among candidates	Trajectory-only model + trajectory + RGB; CNN baseline; nearest baseline	Top-1 accuracy: nearest 30.4%; CNN 39.0%; proposed trajectory 49.0%; proposed trajectory + RGB 62.5% (top-3 92.3%; top-5 97.5%)	Adam optimizer reported; other tools unspecified	Demonstrates value of combining visual cues with spatiotemporal trajectories for tactical prediction.
11	Liu, 2022	Badminton monocular video analytics	TrackNetV2: 77k annotated frames from 26 singles matches; plus 40 extra matches	Hit detection; 3D shuttle trajectory reconstruction	Court geometry, shuttle tracking, player pose, physics priors	HRNet pose; GRU-based shot segmentation; HitNet + optimization; reprojection error	TrackNetV2 dataset; monocular badminton video	Identify hits/segment shots; reconstruct 3D shuttle trajectories	HitNet variants; GRU recurrent network; optimization post-processing	Hit detection: HitNet + optimization recall 94.3%, acc 89.7%, prec 94.9%, F1 0.946; shot segmentation ~90% accuracy	Uses impose for HRNet; other tools unspecified	Illustrates hybrid “deep + physics/domain knowledge” pipeline; notes robustness

			labelled for testing (broadcast-view emphasis)									depends strongly on hit detection.
12	Wu, 2022	Tennis (skill level + stroke classification)	36 right-handed male players (elite/sub-elite/amateur; 12 each)	ITN level discrimination; stroke classification	Wrist IMU (6-axis, 50 Hz); engineered features; PCA	Preprocessing, segmentation, feature extraction, PCA, classification	Single wrist IMU; BLE to smartphone; cloud to computer	Predict stroke type; evaluate skill-level differences	SVM (various scalers), KNN, Naive Bayes; SVM+PCA highlighted	F1 > 0.90 for serve/forehand/backhand and at multiple skill levels; volley lower; ANOVA *p* < .001	SPSS 26 for ANOVA; smartphone + BLE + cloud pipeline reported	Demonstrates feasibility of single-sensor tennis analytics with classical ML and reduced sampling frequency.
13	Yanai, 2022	Basketball decision evaluation (Q-Ball)	619 NBA games (2015–2016 season)	Team/player action value; correlation with team performance; decision alternatives	SportVU trajectories + play-by-play; player attributes; clocks; discrete & continuous actions	Data merging by clock alignment; SHAP for explanation	SportVU tracking + play-by-play	Learn Q-values for actions; recommend alternatives	DRL framework (extension of DDPG handling discrete + continuous); SHAP interpretability [	Correlation: Q-Ball vs points-per-game coefficient 0.622; offensive efficiency vs average points coefficient 0.85 (reported)	unspecified	Connects RL value estimation to coaching “what-if” analysis; notes cold-start and need for prospective evaluation.
14	Costa, 2021	Swimming training analytics (IoT + wearables)	10 athletes (15–17y); ~8000 samples in training setup	Stroke classification; real-time coach feedback	AHRS inertial signals; heart rate; SpO2; turns; derived performance (avg speed, stroke count)	Feature representation over recent positions; ensemble strategy discussion	Wearable AHRS + biosensors (heart rate band; LED SpO2), RF links; pool setup	Classify swim style (6 classes incl. turn)	Random Forest (best), plus mention of NB/KNN/DT/SVM in ensemble framing	Best macro-F1 95.02% (Random Forest); real-time feedback goal	Implemented in Python 3.8.0 (Visual Studio Code)	Emphasizes real-time constraints and trust for coaches; demonstrates high F1 with limited athlete sample.
15	Deliège, 2021	Broadcast soccer understanding (SoccerNet-v2)	500 untrimmed videos; ~300k annotations	Action spotting, camera shot segmentation/boundary detection, replay grounding	ResNet features; temporal context; audio/video cues (AudioVid baseline)	Feature pooling (MaxPool, NetVLAD); context-aware loss (CALF)	SoccerNet videos; broadcast video pipeline	Multi-task benchmark + baselines	NetVLAD/MaxPool, AudioVid, CALF (adaptations)	Action spotting shown/unshown Average-mAP: e.g., NetVLAD (31.4/34.3), AudioVid (39.9/43.0), CALF (40.7/42.1) in excerpted table; Shot segmentation mIoU: basic 35.8 vs CALF 47.3; boundary detection mAP: up to 78.5 (Histogram baseline)	unspecified	Demonstrates multi-task sports broadcast understanding; highlights effect of temporal context and multimodal cues.
16	Gomaz, 2021	Baseball pitching biomechanics (youth)	25 youth pitchers; 10 maximal pitches each	Ball velocity prediction	Peak angular velocity pelvis/trunk; pitcher height	Signal processing to peak angular velocity; multilevel modelling	IMUs on pelvis & sternum (500 Hz); radar gun velocity labels	Predict ball velocity (continuous)	Bayesian regression multilevel models (Full/Personal/Observations)	Full model R^2 0.975, RMSE 0.014; height adds value; reported led comparisons	unspecified	Highlights individualization (height) improves performance; openly references underlying dataset availability.

17	Hauri, 2021	NBA player trajectory prediction	632 NBA games; 113,760 possessions; 1.1M seconds gameplay (0.04s sampling; down sampled)	Future location prediction errors (ADE/FDE)	Player/ball positions and velocities; shot clock; ordered by proximity	Sequence modelling; multi-modal loss; evaluation choosing best mode	NBA tracking dataset (public movement data)	Multi-modal trajectory forecasting for offensive players	Multi-modal LSTM-based models (MBT variants); baselines incl CNN, location-LSTM, SocialGAN	Table reports ADE/FDE (feet) across horizons; e.g., MBT1 for H = 10 ADE 1.43 / FDE 2.98 vs baselines (exact values in table excerpt)	Training on Nvidia GTX 1080; other software unspecified	Demonstrates multi-modality and uncertainty-aware training benefits; highlights horizons and normalization issues.
18	Sen, 2021	Cricket batting shot recognition (video)	Novel dataset CricShot10; also references prior 800-shot dataset; further dataset size details unspecified in excerpt	Multi-class shot classification accuracy	Video frames; temporal dependencies in shot sequences	CNN feature extraction + GRU; transfer learning and fine-tuning (VGG16 variants)	CricShot10 dataset (YouTube-based cricket videos)	Classify 10 batting shot types	VGG16–GRU; transfer learning (VGG16, InceptionV3, Xception, DenseNet169)	VGG16–GRU 86% accuracy; fine-tuned VGG16 variants reach 93% accuracy on CricShot10	unspecified	Demonstrates hybrid CNN-RNN for sports action recognition and the role of transfer learning in small datasets.
19	Luo, 2020	Team-sport valuation (ice hockey)	4.5M NHL play-by-play events (evaluation stated)	Player/action valuation; correlation with success measures; policy-likelihood metrics	NHL play-by-play state/action sequences; sparse goal rewards + learned dense rewards	IRL to infer reward; transfer learning regularization; compute Q-values	NHL play-by-play dataset	Learn dense rewards and rank players/actions	IRL-DK (IRL + domain knowledge), Q-function learning; alternating learning for multi-agent setting	Policy evaluation: NLL/HMD improved (e.g., IRL-DK NLL 49.5; HMD 7.77 vs rule reward NLL 185.0; HMD 13.37); correlations with many success measures reported in tables [19]	unspecified	Motivates IRL for sparse-reward sports; shows denser rewards improve downstream ranking stability.
20	Luu, 2020	NHL injury prediction	2322 players (2007–2017); 2109 skaters + 213 goalies	Next-season injury risk (binary)	Age, **85** performance metrics, injury history (publicly reported)	Feature compilation; model comparison; ensemble	Public injury/performance databases (no official NHL database)	Predict whether injury occurs next season	Random Forest, KNN, Naive Bayes, XGBoost, Top-3 Ensemble; logistic regression baseline	AUC: XGBoost 0.948 (skaters) vs LR 0.937; XGBoost 0.956 (goalies) vs LR 0.947	unspecified	Demonstrates strong discriminative performance but relies on public reporting; highlights ML vs LR comparison.
21	Shao, 2020	Gymnastics video understanding (FineGym)	303 competition records (~708 hours); 530 element categories; 4,883 event instances; 32,697 sub-action instances (v1.0)	Fine-grained action recognition/localization benchmark	Multi-level labels (event/set/element); temporal structure (action/sub-action)	Video annotation hierarchy; baselines include CNN/3D CNN & temporal modelling	FineGym dataset (high-res competition video)	Benchmarking fine-grained recognition & temporal decomposition	Baseline action-recognition pipelines (2D CNN + temporal module; 3D CNN; etc.) described	Paper reports performance still “far from satisfactory” and highlights importance of motion and temporal dynamics	unspecified	Establishes dataset and modelling challenges; open benchmark for fine-grained sports actions.

IRL methods address sparse rewards by inferring dense reward surrogates from demonstrations. Luo et al. (IJCAI 2020) combined Q-function learning with inverse reinforcement learning and domain knowledge (IRL-DK), reporting empirical evaluation based on 4.5M NHL play-by-play events and demonstrating improved policy-likelihood metrics (NLL/HMD) and correlations with multiple success measures (tables provided for offensive and defensive players) (Luo, 2020).

Wearables, IoT, and physiological/biomechanical modelling

Wearables-driven sports ML often confronts smaller n but higher-frequency sensing, requiring careful feature representation and latency-aware design. A swimming analytics framework combined wearable inertial sensing (AHRS) with biosensors (heart rate, pulse oximetry) in a real training setup of 10 athletes (15–17 years) generating ~8000 samples, achieving best macro-F1 95.02% with a Random Forest classifier and implementing the coach-facing tool in Python 3.8 (Costa, 2021).

Tennis stroke classification and skill-level evaluation demonstrate a parallel pipeline: a single wrist-worn IMU (tri-ax accelerometer ±16g; gyroscope ±2000°/s) sampled at 50 Hz and evaluated on 36 right-handed male players, using preprocessing, segmentation, feature extraction, dimensionality reduction (PCA), and classical classifiers (SVM, KNN, Naive Bayes). Reported results highlight SVM configurations achieving F1-scores above 0.90 for several strokes (serve/forehand/backhand) and statistically significant discrimination across skill groups in ITN testing (Wu, et al., 2022).

In sports engineering and biomechanics, ML increasingly integrates physics priors and interpretable feature attribution. A Sports Engineering study trained a multi-output deep neural network to predict final baseball pitch location from ball tracking release metrics and physics-derived context features, using >2 million NCAA Division I pitches; mean Euclidean error on validation was 0.154 m, outperforming linear regression and showing that excluding projectile-motion features degraded performance. Tooling was explicitly reported as Python with scikit-learn 1.4.0 and PyTorch 2.2.2 (Honda, 2022).

For IMU-based prediction of baseball pitch outcomes, Gomaz et al. modelled ball velocity with hierarchical/Bayesian regression using IMUs on pelvis and sternum and radar-gun labels; 25 youth pitchers performed 10 maximal pitches each, and the full model achieved R^2 0.975 with RMSE 0.014 (as reported in their comparison table) (Gomaz, 2021).

Sports medicine and injury risk prediction

Injury risk prediction exemplifies the tension between strong performance claims and practical deployment constraints (imbalance, label noise, causal confounding, and generalization across leagues and seasons). A large NHL study used publicly reported data to predict next-season injury risk, training and comparing multiple ML algorithms (random forest, KNN, Naive Bayes, XGBoost, ensemble) vs logistic regression; XGBoost achieved AUC 0.948 (position players) and 0.956 (goalies) with statistically significant differences vs logistic regression in the reported sample (2007–2017) (Costa, 2021).

Clinical prognosis tasks share similar structure. In adolescent sports-related concussion recovery prediction, sex-stratified ML models were trained using multi-part clinical measures (including VOMS, King-Devick, and C3 Logix Trails); gradient-boosted decision-tree algorithms achieved AUC 0.84 (males) and 0.78 (females) for predicting protracted recovery, improving over statistical baselines (Chu, 2022).

At the team level, soccer injury modelling from training-load deviations combined acute deviations from baseline with ACWR features to classify injury vs baseline epochs, yielding accuracy 0.78, sensitivity 0.73, specificity 0.85 and reporting that SVM-RBF outperformed alternative classifiers for the selected feature subset (Tsilimigkras T. , 2024). A prospective youth soccer screening-battery study (non-elite players) used multiple base classifiers (C4.5, ADTree, SMO, KNN, Random Forest) and imbalance-handling ensembles/resampling, reporting a "*best fit*" model with AUC 0.700 and F-score 0.380 using six field-based measures and emphasizing practical feasibility for screening (Robles-Palazón, 2023). Tableau 1: Representative peer-reviewed studies (2020–2026) on ML in sports.

Quantitative synthesis

Representative study table

Table entries are restricted to publications ≥2020. If an item's peer-reviewed status, software stack, or other detail could not be confirmed from accessible full text/extended abstract, it is explicitly marked unspecified.

Algorithm frequency in the representative table

The following counts are computed from Table 1 by assigning each study a *primary* model family (dataset papers with embedded deep pose components are counted under deep learning/CV). This is a descriptive statistic, not a meta-analytic weight.

Table 1. Descriptive statistic of models.

Algorithm family (Primary)	Count	% (n = 21)
Deep learning	12	57.1%
Tree/boosting ensembles	5	23.8%
Reinforcement learning / IRL	2	9.5%
Classical ML	1	4.8%
Bayesian regression	1	4.8%
Total	**21**	**100%**

Tree-based models remain prominent in injury and clinical prognosis (e.g., NHL injury XGBoost; concussion recovery boosting; swimming RF; youth soccer battery decision-tree ensembles), [3] while deep learning dominates video, pose, and large-scale tracking tasks (SoccerNet-v2/-Tracking, AthletePose3D, PitcherNet) (Deliège, 2021).

Sport/context frequency in the representative table

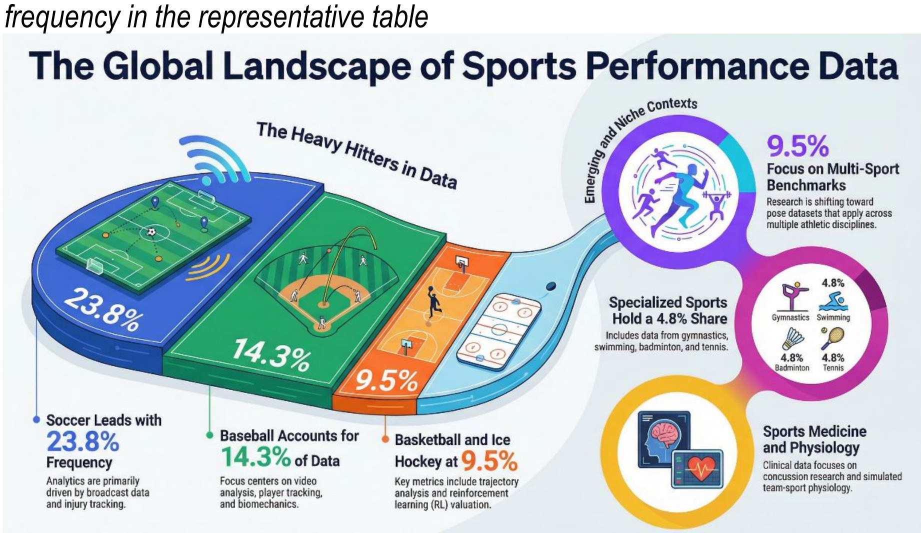

Figure 3. Distribution of machine learning applications across sports domains (2020–2026).

Figure shows a distribution of machine learning applications across sports domains (2020–2026). Soccer represents the largest proportion of published studies (23.8%), driven both by broadcast benchmarks (SoccerNet) and injury-risk modelling studies (Deliège, 2021), followed by baseball (14.3%), baseball representation is amplified by ball tracking and pitch analytics contributions in sports engineering and CV workshops. [11], and basketball and ice hockey (9.5% each). Emerging multi-sport benchmarks account for 9.5%, while specialized sports (e.g., gymnastics, swimming, badminton, tennis) individually represent approximately 4.8%. Sports medicine and physiological applications constitute a distinct and growing research area.

Thematic relationships and timeline

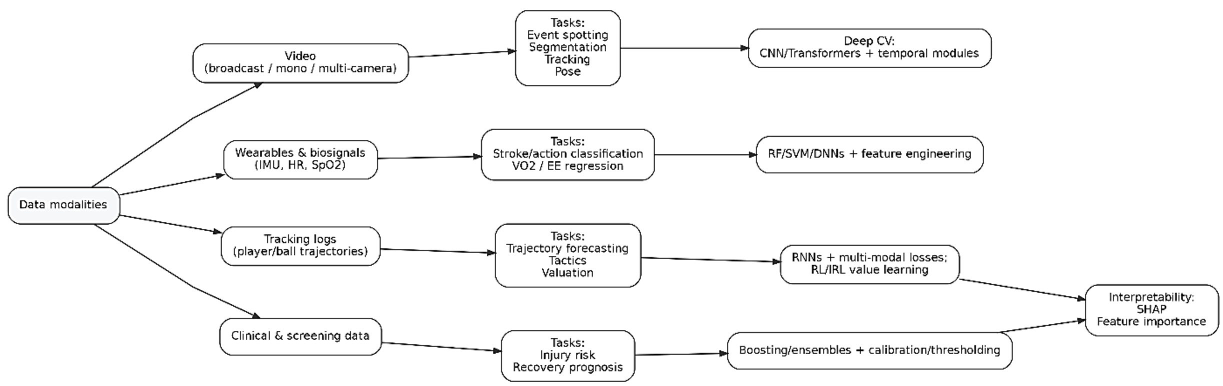

Figure 4. Timeline of representative peer-reviewed ML-in-sports contributions (2020–2025).

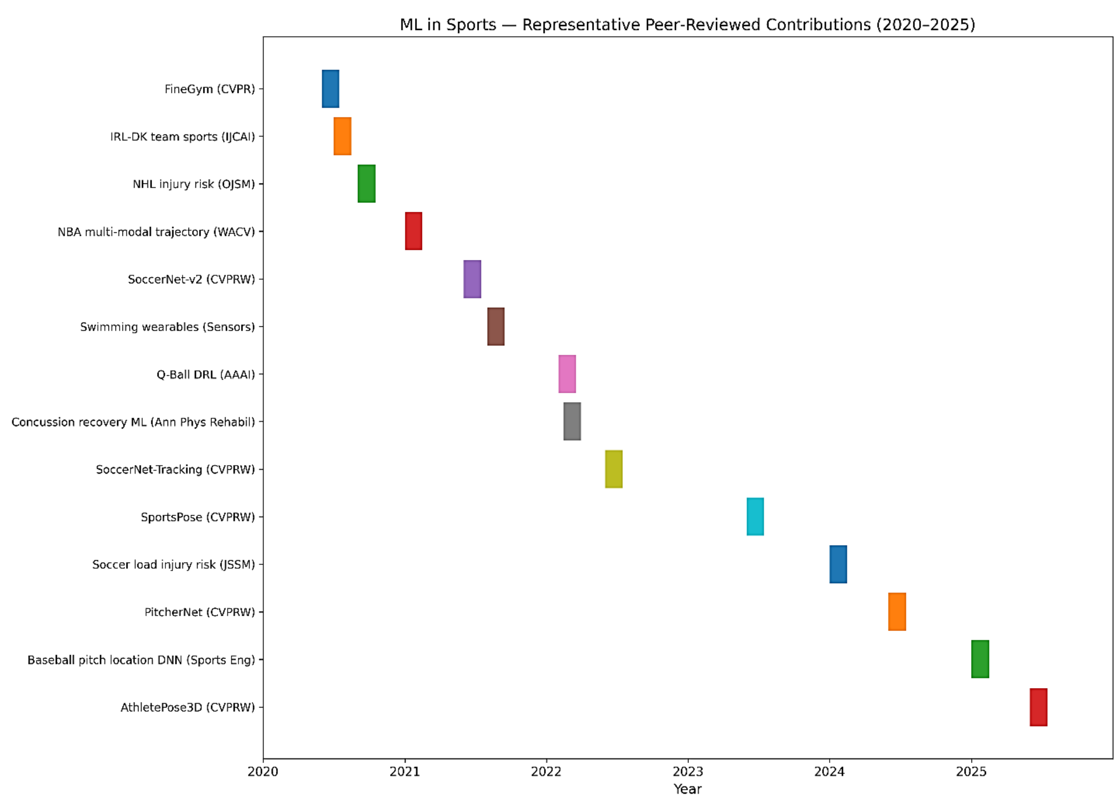

Figure 5. Timeline of representative peer-reviewed machine learning contributions in sports research (2020–2025).

Figure shows a conceptual framework linking data modalities, task families, and dominant machine learning approaches in sports research (2020–2025). Video data primarily supports event spotting, segmentation, tracking, and pose estimation using deep convolutional and transformer-based architectures. Wearable and bio signal data are commonly analysed using classical and ensemble methods for classification and regression tasks. Tracking logs enable trajectory forecasting and reinforcement learning–based valuation, while clinical and screening data are frequently modelled using boosting and calibrated ensemble approaches. Interpretability techniques (e.g., SHAP, feature importance) increasingly support decision translation.

Figure illustrates key benchmark datasets, methodological advances, and applied studies across computer vision, reinforcement learning, wearable analytics, and sports medicine, highlighting the progressive diversification of machine learning applications in sport.

DISCUSSION

The co-evolution of open benchmarks and methodological specialization is a key trend that began in 2020. A common benchmark promotes comparability, identifies failure modes (occlusion, fast motion), and makes domain fine-tuning quantifiable, as demonstrated by SoccerNet-v2 and SoccerNet-Tracking (Cioppa, Soccernet-tracking: Multiple objects tracking dataset and benchmark in soccer videos, 2022). Similarly, "*fine-grained*" sports actions are formalized by FineGym's hierarchical label space, which also reveals that coarse-grained action recognition heuristics do not translate well to subtle, elite motor patterns (Shao, 2020).

The rise of hybrid modelling is a second trend: learning with constraints or priors is often beneficial for sports problems. The monocular badminton pipeline from MonoTrack serves as an example. It combines physics-informed optimization, court geometry, and deep pose and hit detection. The results indicate that upstream error propagation (hit detection) can dominate end-to-end performance (Liu, 2022). Similar to this, the baseball pitch-location study employs projectile predictions derived from physics as contextual features, empirically enhancing deep model performance and confirming that "*learning the physics*" is made simpler when the feature space contains pertinent structure (Honda, 2022).

Third, models that target decisions with downstream consequences make interpretability mandatory. In order to interpret alternative tactical scenarios and feature contributions to Q-values, Q-Ball specifically uses SHAP (Yanai, 2022). To bridge the gap between coach-relevant insight and black-box prediction, permutation-based feature importance is utilized in sports engineering to identify which release metrics influence predicted pitch location (Moore, 2025). The trend toward parsimonious feature subsets in injury-risk contexts—such as six field-based measures in youth screening or seven load-derived features that yield an accuracy of 0.78 in soccer—reflects a desire for deployable screening rules rather than maximal model complexity (Tsilimigkras T., 2024).

Lastly, there is still a clear gap between domains where data are naturally small-n and noisy (physiology, injury events) and domains with massive, labelled data (broadcast video benchmarks, ball tracking). Due to its explicit small sample size ($n = 6$), the PLOS pilot study on VO_2 and energy expenditure is better understood as benchmarking and feasibility in the face of data scarcity rather than as deployable evidence (de Beukelaar, 2025). On the other hand, fields such as NCAA pitch tracking offer millions of examples and can support contemporary deep learning regimes with strong holdout evaluation (Honda, 2022).

CONCLUSIONS

This systematic review consolidates peer-reviewed journal papers and significant conference proceedings published from January 1, 2020, to February 11, 2026, that utilize machine learning (ML) in the domains of sports performance, tactics, officiating/production, athlete monitoring, biomechanics, and sports medicine. The dominant methodological shift since 2020 is the consolidation of deep learning pipelines for vision-based sports understanding (event spotting, tracking, pose estimation), enabled by open benchmarks such as SoccerNet-v2 (broadcast soccer with ~300k annotations across 500 videos) and its extensions for tracking, and fine-grained action benchmarks like FineGym in gymnastics (Deliège, 2021). Deep learning also expanded into trajectory forecasting (e.g., NBA player movement prediction across 632 games and 113,760 possessions) and decision support/reinforcement learning (e.g., Q-value–based basketball evaluation on 619 NBA games) (Hauri, 2021).

Across sports science and medicine, the post-2020 period shows a pragmatic trend toward tree-based ensembles and support vector machines for structured tabular data (training load, screening batteries, injury risk, and clinical/physiological outcomes). Examples include (i) next-season NHL injury prediction with XGBoost achieving AUC ~0.948 for position players and 0.956 for goalies in a dataset of 2322 players (2007–2017) and (ii) adolescent sports-related concussion recovery prognosis where gradient-boosted decision trees achieved AUC 0.84 (males) and 0.78 (females) (Luu, 2020). In soccer injury-risk modelling from training-load deviations, a study combining acute deviations and "*chronic*" (ACWR) features reported accuracy 0.78, sensitivity 0.73, specificity 0.85, with SVM-RBF outperforming alternative classifiers on the selected feature set (Tsilimigkras T. , 2024).

Methodologically, the field increasingly pairs predictive performance with interpretability and actionable insights. This is visible in tactical RL work explicitly using SHAP to interpret recommendations, and in sports engineering work using permutation feature importance to interpret a deep neural network predicting pitch location from >2 million NCAA pitches with mean Euclidean error ~0.154 m and explicit reporting of implementation in Python (scikit-learn, PyTorch) (Yanai, 2022).

From 2020 through early 2026, ML in sports has matured from isolated modelling efforts into an ecosystem structured around open benchmarks (especially for video and pose), richer sequential modelling of multi-agent dynamics, and practical predictive modelling for sports medicine and performance monitoring. Deep learning dominates vision-centric tasks and large tracking datasets, while tree/boosting ensembles and SVM remain central for structured clinical and wearable analytics. Across domains, interpretability methods (SHAP, permutation importance) and hybrid "*learning + priors*" approaches are increasingly used to translate model outputs into coach- and clinician-actionable insights.

The field's next methodological frontier is less about marginal gains on closed benchmarks and more about generalization and deployment: cross-competition domain shift, label quality and missingness, causal confounding in injury and performance studies, and transparent model reporting that supports reproducibility and safe decision-making.

AUTHOR CONTRIBUTIONS

All authors meet the criteria for authorship in accordance with established ethical guidelines. Karim Midoul: conceptualization, methodology, software, data curation, visualization, validation, resources, supervision, writing - review & editing. Badr Eddine EL Mohajir: supervision, validation, writing – review & editing. Outman El Hichami: supervision, validation, writing – review & editing. Adnan Souri: supervision, supervision, validation, writing – review & editing. All authors have critically reviewed and approved the final version of the manuscript and agree to be accountable for all aspects of the work.

FUNDING

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

CONFLICT OF INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this manuscript.

AI USE DISCLOSURE

In accordance with current publishing ethics and transparency recommendations, artificial intelligence (AI) tools were used solely to assist with translation and language editing, with the aim of improving clarity and readability. No AI tools were used in the generation of scientific content, including the study design, data collection, analysis, interpretation of results, or the formulation of conclusions. The authors retain full responsibility for the content of the manuscript and confirm its originality, integrity, and accuracy.

REFERENCES

Bright, J. (2024). Pitchernet: Powering the moneyball evolution in baseball video analytics. IEEE/CVF CVPR. https://doi.org/10.1109/CVPRW63382.2024.00346

Chu, Y. (2022). Machine learning to predict sports-related concussion recovery using clinical data. Annals of Physical and Rehabilitation Medicine, 65(4), 101626. https://doi.org/10.1016/j.rehab.2021.101626

Cioppa, A. (2022). Soccernet-tracking: Multiple objects tracking dataset and benchmark in soccer videos. IEEE/CVF CVPR, (pp. 3491-3502). https://doi.org/10.1109/CVPRW56347.2022.00393

Costa, J. (2021). Framework for intelligent swimming analytics with wearable sensors for stroke classification. Sensors, 21(15), 5162. https://doi.org/10.3390/s21155162

de Beukelaar, T., & Roantree, M. (2025). Estimating oxygen uptake and energy expenditure. PLOS ONE.

Deliège, A. (2021). SoccerNet-v2: A dataset and benchmarks for holistic understanding of broadcast soccer videos. IEEE/CVF CVPR Workshops. https://doi.org/10.1109/CVPRW53098.2021.00508

Gomaz, L. (2021). Individualised ball speed prediction in baseball pitching. Sensors, 21(22), 7442. https://doi.org/10.3390/s21227442

Hauri, S. (2021). Multi-modal trajectory prediction of NBA players. IEEE/CVF WACV. https://doi.org/10.1109/WACV48630.2021.00168

Honda, Y. (2022). Pass receiver prediction in soccer using video and players' trajectories. IEEE/CVF CVPR, (pp. 3503-3512). https://doi.org/10.1109/CVPRW56347.2022.00394

Ingwersen, C. (2023). Sportspose-a dynamic 3d sports pose dataset. IEEE/CVF CVPR, (pp. 5219-5228). https://doi.org/10.1109/CVPRW59228.2023.00550

Liu, P. (2022). MonoTrack: Shuttle trajectory reconstruction from monocular badminton video. IEEE/CVF CVPR, (pp. 3513-3522). https://doi.org/10.1109/CVPRW56347.2022.00395

Luo, Y. (2020). Inverse reinforcement learning for team sports: Valuing actions and players. https://doi.org/10.24963/ijcai.2020/464

Luu, B. (2020). Machine learning outperforms logistic regression analysis to predict next-season NHL player injury: an analysis of 2322 players from 2007 to 2017. Orthopaedic Journal of Sports Medicine, 8(9). https://doi.org/10.1177/2325967120953404

Moore, R. (2025). A context-enhanced deep learning approach to predict baseball pitch location. Sports Engineering, 28, 16. https://doi.org/10.1007/s12283-025-00497-5

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ, 372, n71. https://doi.org/10.1136/bmj.n71

Robles-Palazón, F. (2023). Predicting injury risk using machine learning in male youth soccer players. Chaos, Solitons & Fractals, 167, 113079. https://doi.org/10.1016/j.chaos.2022.113079

Sen, A. (2021). Cricshotclassify: an approach to classifying batting shots from cricket videos using a convolutional neural network and gated recurrent unit. Sensors, 21(8), 2846. https://doi.org/10.3390/s21082846

Shao, D. (2020). FineGym: A hierarchical video dataset for fine-grained action understanding. IEEE/CVF CVPR. https://doi.org/10.1109/CVPR42600.2020.00269

Sheridan, D., Jaspers, A., Cuong, D., Op De Beéck, T., Moyna, N., & de Beukelaar, T. (2025). Estimating oxygen uptake in simulated team sports using machine learning models and wearable sensor data: A pilot study. PLOS ONE, 20(4), e0319760. Récupéré sur https://doi.org/10.1371/journal.pone.0319760

Tsilimigkras, T., Kakkos, I., Matsopoulos, G., & Bogdanis, G. (2024). Enhancing Sports Injury Risk Assessment in Soccer Through Machine Learning and Training Load Analysis. Journal of Sports Science and Medicine, 23, 537-547. https://doi.org/10.52082/jssm.2024.537

Wu, M., Fan, M., Hu, Y., Wang, R., Wang, Y., Li, Y., . . . Xia, G. (2022). A real-time tennis level evaluation and strokes classification system based on the Internet of Things. Internet of Things, 17, 100494. https://doi.org/10.1016/j.iot.2021.100494

Yanai, C. (2022). Q-Ball: Modeling basketball games using deep reinforcement learning. AAAI Conference on Artificial Intelligence, 36, pp. 8806-8813. https://doi.org/10.1609/aaai.v36i8.20861

Yeung, C. (2025). AthletePose3D: A benchmark dataset for 3D human pose estimation and kinematic validation in athletic movements. IEEE/CVF CVPR. https://doi.org/10.1109/CVPRW67362.2025.00592