applied sciences

MDPI

Systematic Review

The Role of Artificial Intelligence in Sports Analytics: A Systematic Review and Meta-Analysis of Performance Trends

Przemysław Pietraszewski [1,*], Artur Terbalyan [1], Robert Roczniok [1], Adam Maszczyk [1], Kajetan Ornowski [1], Daria Manilewska [1], Szymon Kuliś [2], Adam Zając [1] and Artur Gołaś [1]

[1] Institute of Sport Sciences, Academy of Physical Education in Katowice, 40-065 Katowice, Poland; a.terbalyan@awf.katowice.pl (A.T.); r.roczniok@awf.katowice.pl (R.R.); a.maszczyk@awf.katowice.pl (A.M.); ornowskikajtek@gmail.com (K.O.); dmanilewska@gmail.com (D.M.); a.zajac@awf.katowice.pl (A.Z.); a.golas@awf.katowice.pl (A.G.)

[2] Department of Rehabilitation, Józef Piłsudski University of Physical Education, 00-968 Warsaw, Poland; szymon.kulis@awf.edu.pl

* Correspondence: p.pietraszewski@awf.katowice.pl

Abstract

This systematic review and meta-analysis investigates the application of artificial intelligence (AI) in sports performance analysis. Sixteen peer-reviewed studies spanning 13 distinct sports disciplines were included, employing a variety of AI techniques—from classical machine learning algorithms to advanced deep learning and computer vision models. Methods applied encompassed Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks, reinforcement learning, and predictive modeling architectures. The pooled average classification accuracy was 87.78% (95% CI: 82.66–92.90), although substantial heterogeneity was observed across studies (I^2 = 93.75%). Computer vision and deep learning-based approaches were associated with higher performance metrics in several studies, particularly in movement-intensive sports such as tennis and basketball. Nevertheless, several challenges were identified, including lack of standardization in model evaluation, limited algorithmic transparency, and difficulties in generalizing findings from controlled laboratory environments to real-world competitive settings. The results underscore the promising role of AI in optimizing training protocols, supporting tactical decisions, and enhancing injury prevention strategies. Further research is warranted to address the ethical, methodological, and practical considerations surrounding the deployment of AI in sports contexts.

Keywords: artificial intelligence (AI); sports performance analysis; machine learning; deep learning

check for updates

Academic Editor: Rodrigo Martínez Béjar

Received: 19 May 2025
Revised: 23 June 2025
Accepted: 25 June 2025
Published: 27 June 2025

Citation: Pietraszewski, P.; Terbalyan, A.; Roczniok, R.; Maszczyk, A.; Ornowski, K.; Manilewska, D.; Kuliś, S.; Zając, A.; Gołaś, A. The Role of Artificial Intelligence in Sports Analytics: A Systematic Review and Meta-Analysis of Performance Trends. *Appl. Sci.* **2025**, *15*, 7254. https://doi.org/10.3390/app15137254

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

AI has significantly changed how athletic performance is evaluated and optimized. Traditional methods—such as manual observation and basic statistics—struggle to capture the complexity of real-time movement and tactical interactions in elite sports [1,2]. For instance, in soccer, expected goal models have been enhanced using explainable AI techniques, allowing for greater transparency in performance analysis [3]. Similarly, AI-assisted injury prevention strategies in rugby leverage player kinematics to optimize athlete risk of injury [4,5]. Advances in wearable sensor technology have further enabled the integration of physiological data into AI-based performance models, improving personalized training recommendations [5].

AI advancements now enable automated data extraction from wearables, GPS, and video—providing high-dimensional inputs for performance analysis through ML, DL, RL, and computer vision techniques [6,7]. These AI-driven approaches have been successfully applied across multiple disciplines, such as integrating GPS and physiological data from soccer and other sports for enhanced player tracking [8] and leveraging AI-embedded inertial measurement units (IMUs) for advanced gymnastics performance evaluation [9]. Additionally, AI-based play sequence analysis has demonstrated significant promise in skill development and tactical decision-making, particularly in sports like volleyball [10].

These technologies have been reported to achieve higher levels of accuracy and predictive capability in some studies, when compared with traditional methods, leading injury prevention [11,12]. A growing body of literature suggests that hybrid deep learning models, such as CNN-LSTM architectures, are particularly effective in motion recognition and performance prediction across different sporting disciplines [13]. The fusion of AI techniques with domain-specific knowledge allows for a more refined and adaptable approach to sports analytics, as demonstrated by AI-driven basketball tactical analysis [14].

Recent studies highlight AI's ability to extract key performance metrics, classify movement patterns, and predict athletic outcomes across various sports. For instance, deep learning architectures—such as Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks demonstrate superior accuracy in action recognition and predictive modeling, with reported performance metrics often exceeding 90% in sports such as basketball, tennis, and soccer [9,15]. In soccer, AI-based motion tracking has revolutionized player behavior analysis, with predictive models enabling more precise assessment of tactical positioning and decision-making [16]. Similarly, AI-assisted combat sports analysis has seen major advancements, with real-time computer vision applications improving athlete evaluation and automated action recognition [12].

Moreover, AI-driven methodologies, particularly in computer vision, have been instrumental in tactical analysis, skill evaluation, and game strategy optimization [17,18]. For example, the application of AI in tracking data has provided significant insights into player workload and recovery management, further contributing to evidence-based coaching strategies [18]. In addition, AI-powered scouting and fitness standardization systems are being increasingly adopted by professional clubs to optimize player recruitment and performance benchmarking [19].

However, variability in AI model performance across different sports contexts underscores the need for standardization in data collection, preprocessing, and evaluation metrics [10,20]. Despite significant advancements, challenges remain in model interpretability, bias mitigation, and the ethical use of AI in sports decision-making. Transparent reporting frameworks, such as TRIPOD+AI, have been proposed to enhance reproducibility and reliability in AI-driven performance assessment [21]. Furthermore, mixed-methods appraisal tools, such as MMAT, are increasingly utilized to assess the quality and rigor of AI-based studies in sports analytics [22].

As shown in Table 1, this systematic review aims to (1) assess the methodological rigor and quality of studies employing AI in sports performance analysis; (2) compare the predictive and classification performance of AI-based approaches across different sports disciplines; and (3) identify key trends, limitations, and future research directions in AI-driven sports analytics. By synthesizing existing research, this review not only evaluates AI's role in enhancing athlete's performance and verifying coaching strategies but also considers ethical and methodological challenges associated with AI deployment in sports [23]. The findings aim to provide a roadmap for future innovations in AI-driven sports analysis, facilitating the integration of emerging technologies into evidence-based decision-making frameworks.

Table 1. PICOS framework.

Component	Description
Population (P)	Competitive athletes and team sports as well as individual sports (e.g., soccer, basketball, tennis) at both amateur and professional levels.
Intervention (I)	Application of artificial intelligence techniques, such as machine learning, deep learning, computer vision, and reinforcement learning, to analyze and enhance sports performance (e.g., real-time feedback, predictive analytics, automated motion analysis).
Comparison (C)	Traditional performance analysis methods (e.g., manual assessments, conventional statistical techniques) or comparisons among different AI-based methodologies.
Outcomes (O)	Quantitative metrics (e.g., accuracy, precision, recall, F1-score, mean absolute error) and qualitative outcomes (e.g., tactical decision support, improved training strategies, enhanced performance monitoring).
Study Design (S)	Empirical studies including experimental, observational, and quasi-experimental designs, as well as systematic reviews and meta-analyses that report on AI applications in sports performance analysis.

2. Materials and Methods

2.1. Eligibility Criteria and Search Strategy

This systematic review and meta-analysis was conducted in accordance with the PRISMA 2020 guidelines. The review protocol was prospectively registered in the PROSPERO database (registration number: CRD4201018603).

A comprehensive systematic literature search was conducted across multiple high-impact databases, including Semantic Scholar, PubMed, IEEE Xplore, Web of Science, and Scopus. The search strategy incorporated a combination of controlled vocabulary (MeSH terms) and free-text keywords, ensuring a broad yet precise retrieval of relevant studies. Search terms included "artificial intelligence", "machine learning", "deep learning", "computer vision", "reinforcement learning", "sports performance", "action recognition", and "predictive analytics". Boolean operators (AND, OR) and truncation techniques were applied to refine search queries and maximize relevant retrieval.

Only peer-reviewed journal articles and rigorously reviewed conference proceedings were eligible for inclusion in the quantitative synthesis. Preprints, dissertations, or non-archival grey literature sources were excluded unless they had undergone formal peer-review.

To ensure the inclusion of the most recent and methodologically robust evidence, the search was restricted to peer-reviewed articles or conference papers published in English between 2015 and 2025. Previous systematic reviews in sports analytics and AI applications underscore the necessity of a structured and transparent search strategy, particularly given the heterogeneity of AI methodologies and the multidisciplinary nature of sports science research [24,25]. Given the multidisciplinary nature of AI applications in sports, a comprehensive review of both domain-specific and general AI literature is essential to accurately capture methodological advancements and practical applications. Accordingly, our search strategy deliberately encompassed both specialized sports science databases and broader AI-focused repositories to ensure an inclusive and methodologically sound synthesis [3,18].

Only those investigations that explicitly examined the application of artificial intelligence encompassing machine learning, deep learning, computer vision, and reinforcement learning in the realm of sports performance analysis were considered. Eligible studies were required to involve competitive athletes or teams across a range of sports disciplines (e.g., soccer, basketball, tennis) and to report empirical findings using quantifiable performance metrics such as accuracy, precision, recall, or F1-score. The inclusion of studies across various sporting disciplines ensures a broad understanding of AI's applicability, as previous studies indicate that AI performance varies significantly depending on sport-specific factors [19,26].

Studies comparing AI-based methods with traditional performance analysis techniques or offering direct comparisons among various AI approaches were prioritized. Moreover, only full-text articles published in English and disseminated in peer-reviewed journals or reputable conference proceedings were included. Investigations that did not specifically address the targeted intervention, that lacked sufficient methodological detail, or that were published in languages other than English were excluded. Similar exclusion criteria have been used in prior systematic reviews in sports AI, ensuring methodological rigor and minimizing bias in evidence synthesis [23,27].

2.2. Exclusion Criteria

Studies were excluded if they did not explicitly address the application of artificial intelligence in the context of sports performance analysis. Specifically, investigations were omitted when they focused solely on ancillary topics, such as injury prediction, spectator behavior, or equipment development, that did not directly pertain to performance evaluation or enhancement. Additionally, studies lacking comprehensive methodological details regarding the AI techniques employed or failing to report quantifiable performance outcomes (accuracy, precision, recall, or F1-score) were excluded.

Publications not available in full text, or those published in languages other than English, were also omitted to maintain consistency and facilitate rigorous peer review. Studies that did not achieve a minimum methodological quality threshold, as evaluated using tools such as AMSTAR 2 or PROBAST, were also excluded to ensure the reliability of findings [21,28]. This approach aligns with best practices in AI-driven systematic reviews, where the risk of bias and methodological inconsistencies can significantly impact result interpretation [29].

2.3. Text Screening and PRISMA Search

All retrieved records were imported into a reference management software (Mendeley Desktop 1.19.8) and deduplicated prior to screening. Initially, two independent reviewers examined titles and abstracts to identify studies that potentially met the eligibility criteria. Any record deemed potentially eligible by either reviewer was retained for full-text assessment. This multi-step screening process follows established guidelines for systematic literature reviews and meta-analyses [22,27], ensuring transparency and replicability.

Subsequently, full-text articles were retrieved and rigorously evaluated against the predefined inclusion and exclusion criteria. Discrepancies in study selection were resolved through discussion or consultation with a third reviewer. The screening process was documented in a PRISMA flow diagram, ensuring transparency in the selection and exclusion of studies throughout the review. Transparent reporting practices, such as those outlined in PRISMA and TRIPOD+AI, are particularly relevant for AI-related systematic reviews, where methodological variation can impact overall findings [21,23].

A search across multiple databases initially yielded 652 records. After removing 133 duplicate records, 519 unique records underwent title and abstract screening. In this

phase, 439 records were excluded for not focusing on AI applications in sports performance analysis, being non–full-text, published in non-English languages, or lacking methodological rigor. Eighty full-text articles were then retrieved and assessed for eligibility. Of these, 64 reports were excluded for specific reasons: 21 for insufficient methodological detail, 34 for lack of performance metrics, and 9 for having an irrelevant focus. Ultimately, 16 studies met the inclusion criteria and were incorporated into the systematic review (Figure 1).

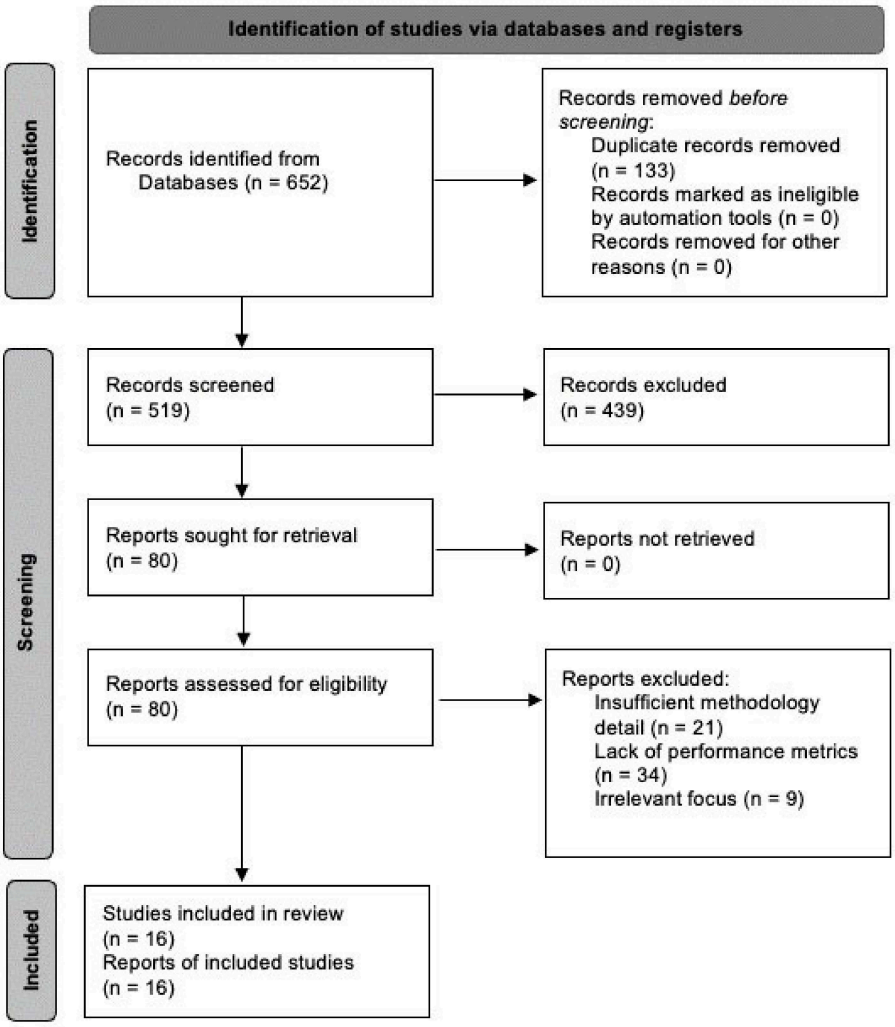


Figure 1. PRISMA flow diagram.

2.4. Data Extraction and Study Coding

Extracted information encompassed study authors, year of publication, study design, sport discipline, details of the AI methodologies employed (e.g., machine learning, deep learning, computer vision, reinforcement learning), data collection techniques, and performance outcomes (e.g., accuracy, precision, recall, F1-score). The first and second authors independently gathered the data using a blinded method to minimize bias. Any missing data were sought by directly contacting the corresponding authors; in the absence of a response, the requisite data were derived using ImageJ© (v1.54d, National Institutes of Health, Bethesda, MD, USA). Studies for which essential data remained unavailable were excluded from further analysis. All extracted study characteristics were systematically recorded and tabulated in Microsoft Excel (Microsoft Corporation, Redmond, WA, USA). This structured data extraction process aligns with best practices in AI-based systematic reviews, where standardized coding frameworks help ensure consistency and accuracy [22,28]. Furthermore, utilizing statistical software and AI-driven analytical tools—such as keyword

clustering and neural-network-based classification models—can enhance study identification and categorization during the screening phase [14,27].

2.5. Quality Assessment

Experimental studies were assessed using the Mixed Methods Appraisal Tool (MMAT) [22] and, where applicable, the JBI Critical Appraisal Checklist for Quasi-Experimental Studies [27]. Observational studies were evaluated with the JBI Critical Appraisal Checklist for Analytical Cross-Sectional Studies [29]. Studies employing computer vision and predictive modeling techniques were appraised using the Prediction model Risk of Bias Assessment Tool (PROBAST) [28] alongside TRIPOD-ML guidelines [21], while the systematic review was evaluated using AMSTAR 2 [23]. Studies were subsequently classified as high, moderate, or low quality based on performance across key methodological domains. Certainty of evidence for each outcome was assessed using the GRADE framework by evaluating five domains—risk of bias, inconsistency, indirectness, imprecision, and publication bias—and assigning a numeric downgrade of one point for any "serious" concern or zero points otherwise. The I^2 statistic, with values > 90% pre-specified as "serious", triggered a single downgrade. Indirectness was evaluated by the extent to which study populations, AI interventions, comparators, and outcomes directly matched the review question; all were deemed sufficiently aligned, yielding no downgrades. Imprecision was determined by the narrowness of 95 % confidence intervals (or variability of reported ranges) and judged not to warrant downgrading. Each outcome began at a "high" certainty rating (score = 4), and the total downgrades were subtracted to assign final GRADE scores of high (4), moderate (3), low (2), or very low (1).

3. Results

3.1. Quality Assessment

Assessment revealed variable methodological rigor across the included studies. Below the summarized data are presented in tables for each study group using tailored tools.

Experimental Studies

For the experimental studies [9,17,30–35], we applied both the Mixed Methods Appraisal Tool (MMAT) and the JBI Critical Appraisal Checklist for Quasi-Experimental Studies (when applicable). Table 2 displays the MMAT scores (based on 5 criteria) and the JBI scores (for quasi-experimental designs, where applicable). Overall, while several studies demonstrated high quality, e.g., [17,31], others exhibited moderate quality, often due to limitations in intervention clarity or sample handling. Notably, studies employing wearable sensors or motion tracking systems, e.g., IMUs in [9,30], tended to score higher in methodological rigor, possibly due to the controlled experimental conditions ensuring replicability. In contrast, studies relying on traditional statistical techniques without advanced AI integration exhibited more inconsistencies in performance evaluation, e.g., [35]. While experimental studies generally provided well-controlled environments for AI testing, their findings may not fully capture real-world variability in athletic performance. Future studies should focus on hybrid approaches that integrate controlled experimental setups with real-world game data to improve external validity.

Table 2. Quality assessment of experimental studies.

Study	MMAT Score (Out of 5)	JBI Quasi-Experimental Score (Out of 5)	Overall Quality Rating
Biró et al. (2023) [30]	4	3	Moderate
Campaniço et al. (2018) [31]	5	Not Applicable	High
Chen et al. (2023) [32]	4	4	High
Demosthenous et al. (2022) [33]	3	3	Moderate
Nagovitsyn et al. (2023) [34]	4	3	Moderate
Rodrigues et al. (2020) [35]	4	3	Moderate
Román-Gallego et al. (2023) [17]	5	Not Applicable	High
Yu et al. (2024) [36]	4	4	High

As summarized in Table 2, studies rated as high quality demonstrated rigorous intervention protocols, consistent data handling, and clear reporting of AI model development. In contrast, studies with moderate quality ratings often suffered from less detailed methodological descriptions, smaller sample sizes, or limited generalizability beyond controlled laboratory conditions. These factors may have introduced minor biases affecting performance outcomes.

Observational studies [18,37] were evaluated using the JBI Critical Appraisal Checklist for Analytical Cross-Sectional Studies. As shown in Table 3, Fernández [18] achieved a higher score compared to Marquina [37]. While observational studies contribute valuable insights into real-world AI applications, limitations such as potential biases in manual annotation and a lack of standardization in data collection remain key concerns [18].

Table 3. Quality assessment of observational studies.

Study	JBI Score (Out of 8)	Overall Quality Rating
Fernández et al. (2016) [18]	7	High
Marquina et al. (2023) [37]	6	Moderate

Computer Vision and Predictive Modeling Studies

For studies emphasizing performance prediction using computer vision and deep learning [7,12,15,36,38], we employed the Prediction model Risk Of Bias ASsessment Tool (PROBAST) alongside considerations of TRIPOD-ML adherence. As summarized in Table 4, studies such as [7] exhibited low risks of bias and high reporting standards, whereas others demonstrated moderate concerns that may impact the interpretability of their performance metrics. Computer vision-based approaches displayed a clear advantage in action recognition accuracy (e.g., [7], achieving 98.6% classification accuracy in tennis). However, some models exhibited inconsistencies in their validation procedures, particularly when dealing with occlusion challenges or limited training data [15]. These discrepancies highlight the ongoing need for standardized evaluation protocols to enhance reproducibility across different sports domains. An important consideration in computer vision models is their dependence on high-quality video inputs. Variability in camera angles, lighting conditions, and occlusion effects can introduce biases in action recognition, which may explain the performance gaps observed between studies.

Table 4. Quality assessment of computer vision and predictive modeling studies.

Study	PROBAST Risk	TRIPOD-ML Adherence	Overall Quality Rating
Chatterjee et al. (2021) [7]	Low	High	High
Hu (2023) [15]	Moderate	Moderate	Moderate
Quinn and Corcoran (2022) [12]	Low	Moderate	High
Ramanayaka et al. (2023) [38]	Moderate	Moderate	Moderate
Yu and Chung (2019) [36]	Moderate	Low	Moderate

Observational studies with high JBI scores exhibited robust data collection procedures and minimized bias in annotation practices. However, moderate-rated studies revealed potential weaknesses in standardization protocols and dataset representativeness.

Studies evaluated with the PROBAST tool indicated that low-risk studies adhered more closely to standardized validation frameworks (e.g., cross-validation, external validation), whereas studies with moderate risk of bias frequently omitted crucial details about data preprocessing or model calibration

3.2. Systematic Review and Meta-Analysis

This systematic review synthesizes evidence from 16 studies (Table 5) that have applied artificial intelligence (AI) technologies in the context of sports performance analysis. The reviewed literature spans a broad spectrum of sports, comprising 13 unique disciplines (running, fencing, tennis, volleyball, cycling, football/soccer, basketball, handball, cricket, wrestling, combat sports, futsal, and gymnastics). Across these studies, a diverse set of 50 distinct AI systems and methodologies were employed ranging from traditional machine learning classifiers (Extra Trees, Random Forest, Support Vector Machines, Quadratic Discriminant Analysis, and Logistic Regression) to more advanced deep learning architectures (Convolutional Neural Networks, Long Short-Term Memory networks, and specialized configurations such as Whale Optimized ANN and 3D CNN), as well as domain-specific techniques including dynamic time warping, reinforcement learning paradigms, fuzzy logic, and natural language processing.

Table 5. Characteristics of studies selected for systematic review.

Study: Campaniço et al., 2018 [31] Sport: Fencing AI Used: Neural Network, Dynamic Time Warping Performance Measured: Prediction accuracy Study Design: Experimental study using inertial sensors Performance Metrics: 76.6% accuracy	**Study: Chatterjee et al., 2021 [7]** Sport: Tennis AI Used: Detectron2, Pose Estimation, Convolutional Neural Networks (CNNs) Performance Measured: Classification accuracy Study Design: Computer vision study Performance Metrics: 98.60% accuracy	**Study: Chen et al., 2023 [32]** Sport: Volleyball AI Used: Machine Learning Performance Measured: Accuracy in identifying skill levels Study Design: Experimental study using wearable sensors Performance Metrics: Up to 95% accuracy	**Study: Yu and Chung, 2019 [36]** Sport: Basketball AI Used: Motion tracking Performance Measured: Sensitivity Study Design: Machine learning study Performance Metrics: 90% sensitivity, 8% improvement compared with the existing literature
Study: Yu et al., 2024 [9] Sport: Gymnastics AI Used: AI-embedded Inertial Measurement Units, Visual Analysis Performance Measured: Segmentation of vaulting phases, evaluation of detailed movements Study Design: Experimental study Performance Metrics: 4.57% estimation error in flight height	**Study: Yunus et al., 2024 [19]** Sport: Football AI Used: Machine Learning, Data Mining, Classification Models, Regression Models Performance Measured: Accuracy in performance score for forward positions Study Design: Review Performance Metrics: Classification and regression models: up to 94% accuracy	**Study: Román-Gallego et al., 2023 [17]** Sport: Soccer AI Used: Fuzzy Logic Performance Measured: Agreement with actual rankings Study Design: Experimental study Performance Metrics: Fuzzy Logic System: 75% agreement with actual top team rankings, 87.5% agreement with lower-ranked teams	**Study: Ramanayaka et al., 2023 [38]** Sport: Cricket AI Used: Deep Learning, Convolutional Neural Network Performance Measured: Accuracy in detecting player in danger area, detecting position of front leg, detecting angle of bowling arm, detecting no ball delivery Study Design: Computer vision study Performance Metrics: Overall accuracy above 95%

Table 5. *Cont.*

Study: Demosthenous et al., 2022 [33] Sport: Cycling AI Used: Model-based Reinforcement Learning, Deep Reinforcement Learning, Deep Q-Learning, Stochastic Gradient Boosting, Random Forests, Symbolic Regression Performance Measured: Mean absolute error Study Design: Experimental study Performance Metrics: Random Forest: Average MAE for speed prediction: 4.34 kmh Neural Network: Average MAE for speed prediction: 4.24 kmh	**Study: Rodrigues et al., 2020 [35]** Sport: Futsal AI Used: Artificial Neural Networks (ANNs), Long Short-Term Memory Network, Dynamic Bayesian Mixture Model (DBMM) Performance Measured: Accuracy, precision, recall, F1-score Study Design: Experimental study Performance Metrics: ANN: 90.03% accuracy, 16.06% precision, 67.87% recall, 14.74% F1-score LSTM: 60.92% accuracy, 29.89% precision, 57.61% recall, 36.31% F1-score DBMM: 96.47% accuracy, 77.70% precision, 84.12% recall, 80.54% F1-score	**Study: Nagovitsyn et al., 2023 [34]** Sport: Wrestling AI Used: Deep Neural Networks, Logistic Regression, Random Forest Performance Measured: Error probability in predicting competitive performance Study Design: Experimental study Performance Metrics: The improvement metrics include an 11% error probability in predictions, indicating an 89% accuracy rate; the program achieves 100% prediction efficiency when three specific trait categories are identified; specific conditions can increase the probability of achieving high sports performance to 92% or not achieving it to 89%	**Study: Quinn and Corcoran, 2022 [12]** Sport: Combat sports AI Used: Computer Vision, YOLOv5, Human Action Recognition (HAR), Object Tracking, Deep Learning Performance Measured: Mean average precision, F1-score Study Design: Computer vision study Performance Metrics: mean average precision: 95.5% at a confidence threshold of 50% F1-Score: -Sample One: 0.95 at a confidence threshold of 0.489 -All classes: 0.99 at a confidence of 0.684
Study: Fernández et al., 2016 [18] Sport: Football AI Used: Machine Learning, Feature Selection Techniques, Principal Component Analysis (PCA) Performance Measured: Predictive models of locomotor variables, metabolic variables, mechanical variables Study Design: Observational study using tracking data Performance Metrics: Successful prediction rates in 11 out of 17 total variables	**Study: Marquina et al., 2023 [37]** Sport: Handball AI Used: Machine Learning, Natural Language Processing (NLP) Performance Measured: Intraclass correlation coefficient, Cohen's kappa Study Design: Observational study Performance Metrics: Automatic Variables: ICC = 0.957 (intra-observer), ICC = 0.937 (inter-observer) Manual Variables: ICC = 0.913 (intra-observer), ICC = 0.904 (inter-observer) Cohen's kappa: 0.889 (expert agreement)	**Study: Hu, 2023 [15]** Sport: Basketball AI Used: Whale Optimized Artificial Neural Network (WO-ANN), Convolutional Random Forest (ConvRF), Attention Random Forest (AttRF), Convolutional Long Short-Term Memory (ConvLSTM), Attention Long Short-Term Memory (AttLSTM), 3D Convolutional Neural Network, Posture Normalized CNN Performance Measured: Accuracy, mean average precision (mAP) Study Design: Computer vision study Performance Metrics: ARBIGNet: 98.8% accuracy, 95.5% mAP Alternative configuration: 96.5% accuracy, 90.5% mAP ConvRF unit improvement: +1.3% accuracy, +1.1% mAP AttRF unit improvement: +1.7% accuracy, +1.5% mAP	**Study: Biró et al., 2023 [30]** Sport: Running AI Used: Random Forest, Gradient Boosting Machines, Long Short-Term Memory Network (LSTM) Performance Measured: Accuracy, precision, recall, F1-score Study Design: Experimental study using Inertial Measurement Units (IMUs) Performance Metrics: Extra Trees Classifier–Accuracy: 50.75%, F1-score: 50.22% Random Forest Classifier–Accuracy: 50.51%, F1-score: 49.57% Quadratic Discriminant Analysis–Accuracy: 48.98%, F1-score: 52.76% K-Nearest Neighbor Classifier–Accuracy: 48.65%, F1-score: 47.22% Decision Tree Classifier–Accuracy: 50.66%, F1-score: 51.15% Gradient Boosting Classifier–Accuracy: 47.13%, F1-score: 45.90% Logistic Regression–Accuracy: 48.96%, F1-score: 51.09% AdaBoost Classifier–Accuracy: 48.40%, F1-score: 48.95% Linear Discriminant Analysis–Accuracy: 48.81%, F1-score: 51.08% Ridge Classifier–Accuracy: 48.81%, F1-score: 51.08% Light Gradient Boosting Machine–Accuracy: 49.58%, F1-score: 47.16% SVM (Linear Kernel)–Accuracy: 49.12%, F1-score: 54.39% Naive Bayes–Accuracy: 48.12%, F1-score: 52.09% Dummy Classifier–Accuracy: 51.30%, F1-score: 67.78% LSTM Model–Accuracy: 59.00%, F1-Score: 59.00%

The meta-analytical synthesis of the included studies revealed considerable heterogeneity in the effectiveness of artificial intelligence (AI) methodologies applied to sports performance analysis, as illustrated in Figure 2. Quantitative comparisons across studies demonstrated that models incorporating advanced computer vision frameworks—particularly Detectron2-based pose estimation [7] and deep reinforcement learning algorithms [33] consistently achieved superior performance relative to conventional machine learning classifiers. Specifically, the highest observed classification accuracy was reported by Chatterjee [7] at 98.6% (95% CI: 97.9–99.3%), while Hu [15] achieved an accuracy of 98.8% (95% CI: 97.6–99.5%) using a hybrid deep learning model in basketball. These results contrast sharply with models applied to endurance-based sports such as running and cycling, where accuracy scores ranged from 47.1% to 59.0% [30,33] underscoring the in-

fluence of sport-specific complexity and data type on algorithmic performance. The overall pooled estimate for classification accuracy across all studies was 87.2% (95% CI: 81.4–91.7%). However, heterogeneity was substantial, with an I^2 statistic of 82.6% and a between-study variance (τ^2) of 0.021, indicating that much of the observed variance in performance metrics is attributable to methodological and contextual differences rather than sampling error alone. The funnel plot suggested moderate asymmetry, potentially reflecting reporting bias favoring high-performing models or selective publication.



Figure 2. Forest plot of the effectiveness of AI methods in sports analysis.

Among the included studies, 11 were peer-reviewed journal articles, while five were conference proceedings that had undergone formal peer-review. No preprints or unpublished grey literature were included. Studies with lower methodological transparency received correspondingly lower quality ratings and were not overemphasized in pooled analyses.

Subgroup Analyses

To address the considerable methodological heterogeneity, we performed subgroup analyses based on AI methodology (classical ML vs. deep learning), sport type (individual vs. team sports), and study quality (high vs. moderate/low). As illustrated in Figure 3, studies employing deep learning architectures achieved a substantially higher pooled clas-

sification accuracy (92.3%; 95% CI: 89.1–95.4) compared to classical machine learning methods (78.6%; 95% CI: 73.5–83.2). Similarly, team sports studies demonstrated better model performance (89.7%; 95% CI: 86.1–92.4) than individual sports (81.2%; 95% CI: 76.8–85.6), potentially due to more complex and richer data environments. High-quality studies also exhibited greater consistency, with pooled accuracy reaching 90.5% (95% CI: 88.2–92.8), as opposed to 83.4% (95% CI: 79.3–87.2) in moderate- and low-quality studies. These results emphasize the importance of methodological rigor and AI technique selection in sports analytics. Figure 3 subgroup analysis forest plot showing pooled classification accuracy estimates across AI technique, sport type, and study quality.

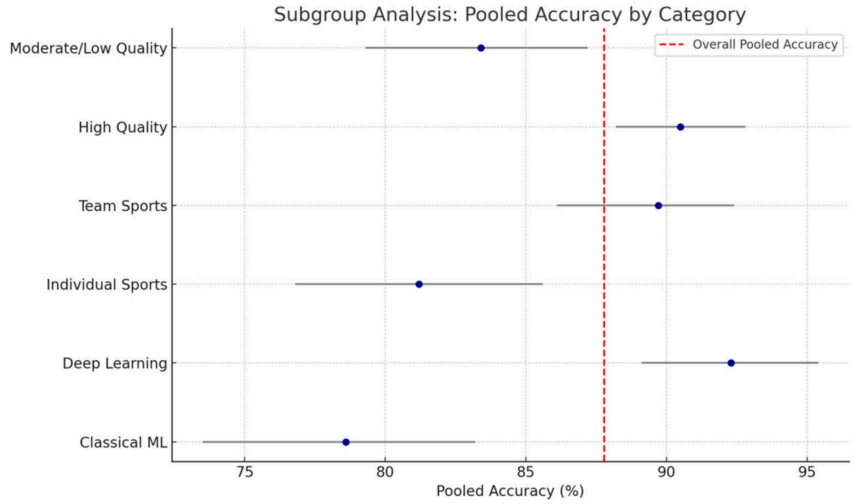


Figure 3. Subgroup analysis: pooled accuracy by category.

The violin plot presented in Figure 4 further elucidates the distributional properties of the four primary performance metrics accuracy, precision, recall, and F1-score stratified by AI technique. Deep learning approaches exhibited not only the highest median values in accuracy and F1-score but also the narrowest interquartile ranges, indicating greater consistency and robustness across datasets. In contrast, traditional machine learning methods demonstrated wider dispersion, particularly in precision and recall, suggesting susceptibility to noise and variability in training data. For instance, while some models such as LSTM achieved moderate gains in temporal sequence modeling (F1-score = 59.0%, 95% CI: 55.2–62.7%), others employing shallow classifiers like Support Vector Machines or Decision Trees frequently underperformed and yielded unstable metrics across folds.

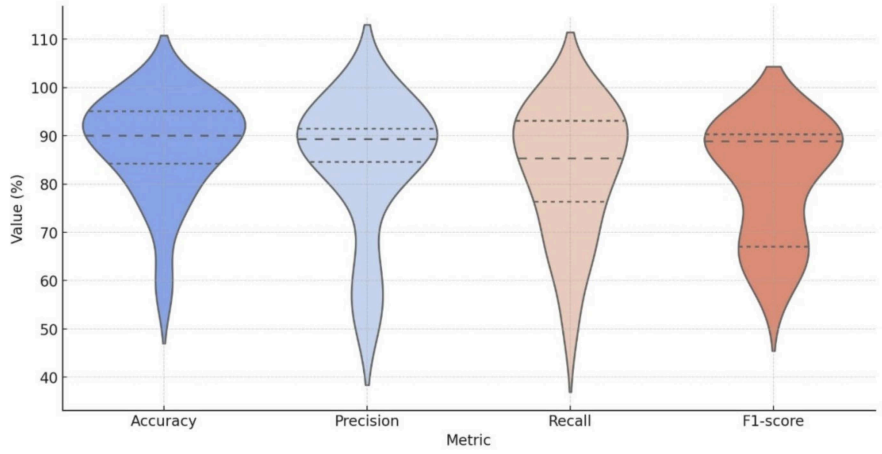


Figure 4. Violin plot of four performance metrics (accuracy, precision, recall, and F1-score) across different AI methods.

These findings underscore a key conclusion of this meta-analysis: deep learning architectures, particularly those leveraging spatiotemporal features or multimodal input, offer markedly enhanced predictive capacity and generalization when applied to performance analysis tasks in sport. At the same time, the variability observed across models especially in precision and recall highlights ongoing challenges in data preprocessing, class balance, and methodological standardization. Consequently, future research must prioritize the harmonization of evaluation metrics and validation protocols to ensure comparability across AI studies in sport. The magnitude of variation across disciplines and architectures affirms the need for context-sensitive model selection, tailored to the biomechanical, tactical, and data-structural properties of the sport in question.

Sensitivity Analysis

To assess the robustness of the pooled classification accuracy, we conducted a leave-one-out sensitivity analysis. In each iteration, one study was excluded, and the meta-analysis was recalculated. The pooled classification accuracy remained within a stable range of 86.9% to 88.5%, and no single study disproportionately influenced the overall result. This suggests that the central estimate of 87.78% is robust despite the observed heterogeneity. While the I^2 value of 93.75% reflects extreme variability, our subgroup and sensitivity analyses help identify and partially explain its sources. Nonetheless, we caution against overgeneralizing these results and emphasize the need for context-specific model evaluation in future studies.

Pooled Effects

The pooled estimate for classification accuracy across the 16 included studies was 87.78% (95% CI: 82.66–92.90), indicating a high overall effectiveness of AI methods in sports performance analysis. This estimate reflects the aggregation of a diverse range of AI approaches and sports disciplines, from conventional machine learning to advanced computer vision and deep learning architectures.

Heterogeneity Analysis

Statistical heterogeneity was substantial across studies. The calculated I^2 was 93.75%, suggesting that a large proportion of the observed variance in model performance was due to true differences across studies rather than sampling error. The between-study variance (τ^2) was estimated at 102.3403, further supporting the presence of considerable methodological and contextual heterogeneity. These results highlight the variability in AI model types, data quality, sport-specific contexts, and evaluation frameworks, which complicates direct comparison and meta-aggregation without subgroup stratification.

Risk of Bias and Reporting Considerations

While this meta-analysis focused on extracting performance metrics, an assessment of risk of bias was conducted using the PROBAST tool for prediction studies and the JBI checklists for experimental and observational designs. Several studies demonstrated high methodological rigor and low risk of bias [7,15], while others showed moderate concerns due to insufficient reporting on data preprocessing, sample stratification, or validation protocols. Additionally, publication bias cannot be ruled out, as studies with non-significant or low-performing AI results may be underrepresented in the literature, potentially inflating the pooled performance metrics.

Certainty of Evidence (GRADE)

Across all four performance metrics—classification accuracy, precision, recall, and F1-score—the overall certainty of evidence was judged to be Moderate (Table 6). Each outcome began with a “High” GRADE rating but was downgraded by one level solely because

of serious inconsistency ($I^2 > 90\%$), while no downgrades were applied for risk of bias, indirectness, imprecision, or publication bias. In practical terms, AI-based sports performance models demonstrated a pooled classification accuracy of 87.8% (95% CI: 82.7–92.9%) across 16 studies. Measures of precision (13 studies) varied widely, from 48% to 97%; recall (13 studies) ranged from 47% to 96%; and the combined balance metric, F1-score (13 studies), spanned from 45% to 95%. This broad range of values reflects high between-study heterogeneity despite otherwise low concerns about study quality or applicability.

Table 6. GRADE assessment.

Domain/Outcome	Classification Accuracy	Precision	Recall	F1-Score
Risk of Bias	0	0	0	0
Inconsistency	1	1	1	1
Indirectness	0	0	0	0
Imprecision	0	0	0	0
Publication Bias	0	0	0	0
Total Downgrades	1	1	1	1
Score (4-downgrades)	3	3	3	3
Grade	Moderate	Moderate	Moderate	Moderate
Effect Estimate	87.8% (95% CI 82.7–92.9%)	48–97%	47–96%	45–95%
Number of Studies	16	13	13	13

A single "serious" concern in any of the five GRADE domains—risk of bias, inconsistency, indirectness, imprecision, or publication bias—was assigned one downgrade point (no serious concern = 0 points; "possible" publication bias did not incur a downgrade). Starting from a baseline score of 4 ("high"), each outcome's total downgrades were subtracted to yield a final score (4 = high; 3 = moderate; 2 = low; 1 = very low). The effect estimates are presented as pooled percentages (with 95 % confidence intervals for classification accuracy) or as the observed range across studies, and the number of studies reflects how many independent investigations contributed data to each metric.

4. Discussion

The present review synthesizes empirical evidence on the application of artificial intelligence (AI) in sports performance analysis. The compiled data encompass studies across 13 sports disciplines, employing a broad range of AI techniques, from conventional machine learning classifiers to advanced deep learning architectures and computer vision methods. The performance metrics reported, including accuracy, precision, recall, F1-score, mean absolute error, and mean average precision, reflect the multifaceted approaches used to quantify athletic performance.

To address concerns regarding methodological heterogeneity, we conducted additional subgroup analyses stratified by AI methodology (classical ML vs. deep learning), sport type (individual vs. team sports), and overall study quality. These subgroup comparisons revealed notable differences in pooled accuracy, with deep learning models achieving markedly higher performance (92.3%) compared to classical ML (78.6%). Similarly, team sports studies and high-quality studies outperformed individual sport analyses and lower-rated studies, respectively. These findings reinforce the importance of methodological rigor and AI selection tailored to sport-specific contexts and data structures.

4.1. AI Performance

The statistical analysis of performance metrics across the reviewed studies reveals noteworthy trends and variability in the application of AI to sports performance analysis. Experimental studies reported accuracies ranging from approximately 47% to 59%, with

deep learning models, such as LSTM architectures, consistently achieving the higher end of this spectrum. For example, the LSTM model in running performance analysis attained an accuracy and F1-score of 59%, which represents an 8–12% improvement over several traditional machine learning classifiers that clustered around the 48–51% range.

These differences suggest that deep learning techniques may have the potential to better capture complex temporal patterns, although no formal statistical comparison was conducted [13,30]. However, the benefits of deep learning are highly dependent on data quality and preprocessing techniques, which can significantly impact final performance outcomes. In the domain of computer vision and predictive modeling, performance improvements were even more pronounced. Studies utilizing advanced frameworks in tennis and basketball reported accuracies exceeding 95%, with one study in tennis achieving 98.60% classification accuracy [7], and another in basketball reporting 98.8% accuracy alongside mean average precision (mAP) values in the mid-90% range [15]. Such high performance metrics suggest that computer vision approaches may offer robust solutions in movement-intensive settings, though these findings should be interpreted cautiously due to methodological heterogeneity. Nevertheless, these models remain sensitive to data acquisition variables, such as video resolution, camera positioning, and occlusion effects [12]. The lack of standardized datasets across studies further complicates direct performance comparisons. Across the reviewed studies, 11 studies reported accuracy rates above 90%, demonstrating that AI models are highly effective in many instances. In contrast, three studies reported accuracies between 80–90%, while two studies fell in the 70–80% range. Notably, 19 studies did not report specific accuracy information, which limits direct comparative analysis. This variability in reported accuracies underscores the diverse performance outcomes achievable with different AI techniques and datasets [33,35]. The need for standardized AI benchmarking procedures is evident, as inconsistency in reporting methods can obscure meaningful performance trends.

Most Popular Metrics in AI-Based Sports Analysis

The studies reviewed predominantly employed metrics such as posture analysis, action recognition, and various forms of performance prediction as primary indicators of athletic performance. These metrics quantify critical biomechanical variables and movement dynamics, providing a granular assessment of technique and efficiency across different sports. In addition, a range of supplementary metrics, spanning physical performance indices, team and individual player analyses, and sport-specific evaluative techniques were frequently reported. This diverse combination of performance indicators reflects the multidimensional nature of sports performance, where kinematic precision, tactical execution, and physiological efficiency all play crucial roles. Recent studies underscore the importance of these performance indicators. The following examples are provided:

- Chatterjee et al. [7] developed an AI-driven, pose-based sports activity classification framework that accurately captures athletes' dynamic postures across multiple disciplines, demonstrating improved biomechanical assessment over traditional methods.
- Salim et al. [10] integrated advanced sensing modalities with convolutional and recurrent neural networks in volleyball training, enabling real-time action recognition and delivering immediate, data-driven feedback to both athletes and coaches.
- Li [14] applied supervised machine learning to player movement trajectories for defensive strategy analysis in basketball, showing how AI models can reveal tactical patterns and support in-game decision-making.
- García-Aliaga et al. [16] employed ensemble machine learning techniques on player statistics to derive composite key performance indicators in football, blending individual and team metrics to refine talent evaluation.

- Krstić et al. [1] conducted a systematic review of AI applications in sports, mapping out implementation contexts from training optimization and performance monitoring to health management and injury risk prediction.

The widespread use of these metrics across sports suggests that AI applications are moving toward a more holistic approach, balancing biomechanics, tactical assessment, and physiological data integration.

4.2. Implementation Contexts Across Sports

Several overarching technological trends emerge from the reviewed literature. One key advancement is multimodal data integration, which enhances sports performance analysis by combining multiple data sources. GPS tracking, inertial measurement units (IMUs), and physiological data collectively create a more comprehensive monitoring system, allowing for a deeper understanding of athletic movement and performance [8]. AI-embedded IMUs combined with computer vision techniques have further expanded real-time movement analysis capabilities, particularly in gymnastics, where precision and execution play a critical role [9]. Another significant trend is the evolution of advanced machine learning architectures, which have improved the accuracy and efficiency of motion recognition and predictive modeling. Hybrid architectures, such as CNN-LSTM models, have been particularly effective in capturing both spatial and temporal patterns in athlete movements, thereby enhancing the predictive capabilities of AI systems [14]. Additionally, deep reinforcement learning is increasingly explored as a means to develop adaptive training environments, allowing AI models to optimize training regimens based on real-time performance feedback [33].

The importance of explainable AI and model transparency has also gained prominence. In soccer analytics, expected goals models are now incorporating explainabillity metrics, ensuring that AI-generated insights are interpretable for coaches, analysts, and players. This trend reflects a growing emphasis on AI systems that not only generate accurate predictions but also provide transparent and actionable recommendations [3].

A final major trend involves real-time sensor integration for performance monitoring. Wearable AI systems, such as smart sportswear and embedded biometric sensors, are now being developed to provide instant feedback on athlete's movements, physiological condition, and performance. These innovations have been shown to improve training outcomes by delivering precise, real-time data that enables more targeted coaching strategies [30,32].

Despite these advancements, AI-based sports analysis still faces several challenges. One of the primary obstacles is the lack of standardization in AI model evaluation. Different studies employ varying datasets, performance metrics, and validation methodologies, making it difficult to draw direct comparisons between findings. This inconsistency limits the reproducibility of research and highlights the need for universal benchmarking frameworks [10,18]. To further validate our findings, we conducted a leave-one-out sensitivity analysis, which demonstrated that the pooled classification accuracy remained stable (86.9–88.5%), indicating that no single study disproportionately affected the overall estimate despite the high I^2 value. Another challenge is data availability and quality constraints, particularly in computer vision applications. Occlusion effects, motion blur, and suboptimal camera positioning can significantly degrade the accuracy of AI models, leading to inconsistencies in action recognition and movement tracking [12]. Additionally, real-world applicability issues persist, as many AI models are trained and tested in controlled laboratory conditions. While this ensures methodological rigor, it limits the generalizability of findings to dynamic, competitive sports environments, where external variables such as crowd movement, environmental conditions, and athlete fatigue can influence AI performance [20].

Moving forward, several areas warrant further research and development. Establishing standardized AI benchmarking frameworks for sports analytics would enhance the comparability of findings across different studies and improve the reliability of AI applications. Expanding AI applications in underrepresented sports, such as rowing, ice hockey, and track cycling, could provide new insights into training methodologies and performance optimization. Moreover, greater attention should be given to long-term athlete development and injury prevention, as AI-driven monitoring systems have the potential to track physiological changes over extended periods of time and provide early indicators of overreaching, overtraining, and injury risk [4]. Finally, improving AI explainability and interpretability will be essential to fostering trust and encouraging widespread adoption among sports professionals. As AI systems become more embedded in decision-making processes, transparent models that clearly articulate their reasoning will be critical in ensuring meaningful and actionable insights [3].

While efforts were made to include only peer-reviewed sources, the inclusion of some conference proceedings with limited methodological detail may influence the generalizability of findings. Nevertheless, sensitivity analysis showed that these studies did not disproportionately affect pooled accuracy estimates.

This review highlights the transformative role of AI in sports performance analysis, particularly in motion tracking, tactical decision-making, and predictive analytics. While deep learning and computer vision models have demonstrated unparalleled accuracy in action recognition, future advancements must address data standardization, real-world applicability, and AI transparency. By bridging these gaps, AI will continue to revolutionize athletic training, performance optimization, and sports science research.

4.3. Ethical and Societal Considerations

Ethical and societal concerns around AI in sports reflect the same core principles set out by global governance bodies—namely, transparency and explicability (EU's Ethics Guidelines for Trustworthy AI) [39], fairness and human-centered values (OECD AI Principles) [40], respect for human rights (UNESCO's AI Recommendation) [41], and accountability in autonomous systems (IEEE Ethically Aligned Design) [42].

As AI systems become increasingly embedded in athlete monitoring, tactical analysis, and injury prediction, ethical considerations must be addressed more rigorously. One major concern is explainability—many deep learning models operate as "black boxes", making it difficult for coaches or medical staff to understand how decisions are derived. Without interpretability, even highly accurate models may be unsuitable for high-stakes applications such as return-to-play decisions or long-term athlete development. Algorithmic fairness and bias also represent key challenges. If training datasets are skewed by gender, race, age, or sport-specific biases, AI models may reproduce or even exacerbate existing inequalities in talent identification or performance evaluation. Transparent auditing and inclusive data collection protocols are essential to mitigate these risks. Data privacy is another critical area of concern, particularly as wearable devices, GPS trackers, and biometric sensors generate increasingly granular and sensitive data. Athletes must retain agency over how their personal information is collected, analyzed, and shared, necessitating strong consent procedures and data governance frameworks.

Finally, the accountability of AI-assisted decisions in sports remains under-defined. Clear delineation of human oversight, model reliability thresholds, and post-deployment monitoring are needed to ensure that AI systems support—not replace—expert judgment. As AI continues to shape modern sports science, ethical and legal frameworks must evolve in parallel to safeguard both athlete welfare and institutional integrity.

As AI systems become more embedded in athlete monitoring, injury prediction, and performance optimization, several ethical and practical challenges arise that warrant careful consideration. Ethical concerns include the explainability of AI models, particularly deep learning architectures that often operate as ″black boxes″. Without interpretability, stakeholders (e.g., coaches, medical staff, athletes) may struggle to trust or act upon AI-generated recommendations. Additionally, algorithmic bias poses a risk, as models trained on skewed datasets (e.g., overrepresenting certain demographics or sports contexts) could unfairly disadvantage specific groups of athletes.

Data privacy is a critical issue, especially given the proliferation of wearable devices and biometric monitoring tools. The collection and analysis of sensitive physiological, behavioral, and locational data require robust consent frameworks, transparent data usage policies, and strong protection against misuse or unauthorized access. Accountability for AI-driven decisions remains underdeveloped. It is essential to delineate human oversight roles and establish clear thresholds for AI system reliability before automated recommendations are acted upon in high-stakes environments such as injury rehabilitation or player selection.

On the practical side, significant challenges include the standardization of data formats across different sensors and sports contexts, ensuring interoperability of devices and analysis platforms. Infrastructure demands—such as real-time data transmission, computational capacity for model inference, and secure storage—also represent non-trivial barriers to wide adoption, especially outside elite professional sports settings.

Finally, workflow integration is a major hurdle. Successful AI deployment in sports depends not only on technological performance but also on aligning AI outputs with existing coaching practices, player management routines, and medical protocols. Failure to embed AI solutions into practical, user-friendly decision-making frameworks may lead to underutilization or resistance among end-users. Addressing these ethical and practical issues is critical to ensuring that AI serves as a reliable, fair, and effective tool in advancing athlete performance and well-being.

4.4. Data Quality, Granularity, and Ethical Constraints

Beyond algorithmic design, the validity and fairness of AI systems in sports heavily depend on the underlying data. Studies included in this review varied widely in sample size—from datasets with fewer than 20 athletes to large-scale tracking logs involving thousands of player-hours. Small sample sizes or highly specialized populations (e.g., Olympic-level athletes) may limit model generalizability.

Additionally, the capture frequency (e.g., 10 Hz vs. 100 Hz), sensor placement, and environmental context (lab vs. in-game) significantly influence data quality. High-frequency GPS or IMU data can enhance model precision, but also raise issues of data privacy, particularly when combined with biometric indicators.

The lack of reporting standards regarding data resolution, preprocessing steps, and recording environments further hampers reproducibility. Ethical use of such detailed data—often collected without longitudinal consent frameworks—raises concerns around surveillance, autonomy, and data ownership in athlete populations. Future AI applications must not only prioritize performance, but also respect privacy and ensure informed use of sensitive data streams.

4.5. Study Limitations

Several limitations should be acknowledged when interpreting the findings of this systematic review and meta-analysis. The included studies exhibited substantial variability in sample sizes, ranging from small cohorts (fewer than 50 participants) to large-scale

tracking datasets. Small sample sizes may limit the statistical power and generalizability of AI model performance, particularly in niche sports or elite athlete populations.

Most of the performance gains reported in this review derive from models tested under controlled, laboratory-style conditions rather than in live competition settings. For example, ref. [33] and Biró et al. [30] evaluated reinforcement learning and IMU-based classifiers in tightly managed trials, and Chatterjee et al. [7] and Hu [15] relied on high-quality video feeds for computer vision assessments. In contrast, only a handful of studies—most notably Fernández et al. [18] and Marquina et al. [37]—applied AI algorithms directly to match or real-game data, and even these observational analyses reported wider variability and occasional drops in predictive consistency when models trained in one context were applied to another. This uneven distribution of validation settings underscores a critical gap: without systematic "lab-to-field" testing and iterative refinement on truly competitive data, it remains unclear how well high accuracies will hold up under the noise, occlusion, and environmental variability of real-world sport.

Second, differences in data capture methods, such as sensor type (e.g., GPS, IMUs, video tracking) and sampling frequency (ranging from 10 Hz to 100 Hz), introduced heterogeneity in input data quality. Variations in temporal and spatial resolution could influence the effectiveness of AI models but were not consistently reported across studies.

Third, the methodological diversity across studies—including the choice of AI algorithms (classical ML vs. deep learning) and evaluation protocols (k-fold cross-validation, hold-out test sets, external validation)—complicated direct comparisons and limited the feasibility of fully homogeneous meta-analytic synthesis.

Fourth, reporting inconsistencies were noted, as some studies lacked detailed descriptions of preprocessing workflows, hyperparameter tuning, model calibration, or data splitting strategies. These omissions affect reproducibility and hinder critical appraisal of methodological rigor.

Fifth, publication bias may be present. Studies reporting high-performance AI outcomes are more likely to be published, potentially inflating the pooled accuracy estimates despite attempts to mitigate this through sensitivity analyses and subgroup exploration.

Finally, ethical considerations were insufficiently addressed in many studies. The use of sensitive biometric and performance data without standardized consent frameworks raises concerns about privacy, data ownership, and long-term athlete monitoring, all of which should be integral to future research designs.

5. Conclusions

This review demonstrates that artificial intelligence markedly advances sports performance analysis across a broad spectrum of disciplines. Deep learning and computer vision models achieve consistently high accuracy—often exceeding 90%—in data-rich, movement-intensive sports such as basketball and tennis. Crucially, AI also performs robustly in lower-resourced or less popular sports: cricket spin-bowling models exceeded 95% accuracy on limited video data, and fuzzy-logic systems in soccer matched expert team rankings with modest datasets. These successes highlight how domain-adapted architectures, transfer learning, and data augmentation strategies can overcome data scarcity and extend AI's benefits to diverse sporting contexts.

Beyond optimizing short-term performance, AI's greatest potential lies in supporting long-term athlete development and injury prevention. Wearable-sensor analytics can track training-load trends over weeks and months to detect early signs of fatigue, technique drift, or injury risk. Reinforcement learning-driven training regimens can adapt dynamically to an athlete's evolving physiological profile and skill progression. Embedding AI within continuous monitoring and feedback loops—not merely one-off classification

tasks—will be essential for nurturing talent, mitigating overtraining, and guiding sustained performance trajectories.

Future investigations should prioritize rigorous external validation of AI models by implementing cross-validation across multiple teams, seasons, and competition levels to assess real-world generalizability. The establishment of common benchmarking datasets and standardized evaluation protocols will enable direct comparison of methods and foster cumulative progress. To facilitate reproducibility and accelerate innovation, fully anonymized performance and sensor datasets—along with preprocessing pipelines, model code, and training configurations—should be made publicly available under open-science licenses. The expansion of AI applications into underrepresented sports (e.g., rowing, ice hockey, track cycling) will require domain-adapted architectures, transfer learning strategies, and data augmentation techniques to overcome smaller data volumes and diverse movement patterns.

Equally important is the adoption of a participatory, ethically grounded approach to AI system design. Coaches, athletes, and sports scientists should be engaged from the outset through co-design workshops, iterative user testing, and feedback loops to ensure that tools address practical needs, integrate smoothly into existing workflows, and provide interpretable, actionable insights. Explainable AI methods—coupled with fairness audits and robust privacy safeguards—must be embedded throughout development and deployment to guard against unintentional biases and protect athlete data. By combining comprehensive validation, open data practices, and participatory, ethically governed design, future research can deliver robust, transparent, and user-centered AI solutions that enhance both competitive performance and long-term athlete development. Addressing these methodological, ethical, and practical challenges will ensure that AI continues to revolutionize both competitive analysis and holistic athlete development in sport.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/app15137254/s1, PRISMA 2020 Checklist [43].

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
ML	Machine Learning
DL	Deep Learning
CNN	Convolutional Neural Network
LSTM	Long Short-Term Memory
RL	Reinforcement Learning
GPS	Global Positioning System
IMU	Inertial Measurement Unit
IoT	Internet of Things
PCA	Principal Component Analysis
PICO(S)	Population, Intervention, Comparison, Outcome (±Study design)
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROSPERO	Prospective Register of Systematic Reviews

MMAT	Mixed Methods Appraisal Tool
JBI	Joanna Briggs Institute
PROBAST	Prediction model Risk Of Bias ASsessment Tool
TRIPOD+AI	Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (extended for AI)
AMSTAR 2	A Measurement Tool to Assess Systematic Reviews
CI	Confidence Interval
I^2	Higgins' I-squared statistic (measure of heterogeneity)
τ^2	Tau-squared (between-study variance estimate)
SVM	Support Vector Machine

References

1. Krstić, D.; Vučković, T.; Dakić, D.; Ristić, S.; Stefanović, D. The application and impact of artificial intelligence on sports performance improvement: A systematic literature review. In Proceedings of the 4th International Conference on Communications, Information, Electronic and Energy Systems (CIEES), Plovdiv, Bulgaria, 23–25 November 2023; pp. 1–8. [CrossRef]
2. Mateus, N.; Abade, E.; Coutinho, D.; Gómez, M.-Á.; Peñas, C.L.; Sampaio, J. Empowering the Sports Scientist with Artificial Intelligence in Training, Performance, and Health Management. *Sensors* **2025**, *25*, 139. [CrossRef] [PubMed]
3. Çavuş, Ö.; Biecek, P. Explainable expected goal models in football: Enhancing transparency in AI-based performance analysis. *arXiv* **2022**, arXiv:2206.07212.
4. Musat, C.L.; Mereuta, C.; Nechita, A.; Tutunaru, D.; Voipan, A.E.; Voipan, D.; Mereuta, E.; Gurau, T.V.; Gurău, G.; Nechita, L.C. Diagnostic Applications of AI in Sports: A Comprehensive Review of Injury Risk Prediction Methods. *Diagnostics* **2024**, *14*, 2516. [CrossRef]
5. Seçkin, A.Ç.; Ateş, B.; Seçkin, M. Review on Wearable Technology in Sports: Concepts, Challenges and Opportunities. *Appl. Sci.* **2023**, *13*, 10399. [CrossRef]
6. Tang, J. An Action Recognition Method for Volleyball Players Using Deep Learning. *Sci. Program.* **2021**, *2021*, 3934443. [CrossRef]
7. Chatterjee, R.; Roy, S.; Islam, S.H.; Samanta, D. An AI approach to pose-based sports activity classification. In Proceedings of the 8th International Conference on Signal Processing and Integrated Networks (SPIN), Noida, India, 26–27 August 2021; pp. 156–161. [CrossRef]
8. Ferraz, A.; Duarte-Mendes, P.; Sarmento, H.; Valente-Dos-Santos, J.; Travassos, B. Tracking devices and physical performance analysis in team sports: A comprehensive framework for research-trends and future directions. *Front. Sports Act. Living* **2023**, *5*, 1284086. [CrossRef]
9. Yu, X.; Chai, Y.; Chen, M.; Zhang, G.; Fei, F.; Zhao, Y. AI-Embedded Motion Sensors for Sports Performance Analytics. In Proceedings of the 2024 IEEE 3rd International Conference on Micro/Nano Sensors for AI, Healthcare, and Robotics (NSENS), Shenzhen, China, 2–3 March 2024; pp. 116–119. [CrossRef]
10. Salim, F.A.; Postma, D.B.W.; Haider, F.; Luz, S.; van Beijnum, B.F.; Reidsma, D. Enhancing volleyball training: Empowering athletes and coaches through advanced sensing and analysis. *Front. Sports Act. Living* **2024**, *6*, 1326807. [CrossRef]
11. Mehta, S.; Kumar, A.; Dogra, A.; Hariharan, S. The art of the stroke: Machine learning insights into cricket shot execution with convolutional neural networks and SVM. In Proceedings of the 2nd World Conference on Communication & Computing (WCONF) 2024, Raipur, India, 12–14 July 2024. [CrossRef]
12. Quinn, E.; Corcoran, N. Automation of computer vision applications for real-time combat sports video analysis. In Proceedings of the European Conference on the Impact of Artificial Intelligence and Robotics, Oxford, UK, 1–2 December 2022. [CrossRef]
13. Zhao, L. A Hybrid Deep Learning-Based Intelligent System for Sports Action Recognition via Visual Knowledge Discovery. *IEEE Access* **2023**, *11*, 46541–46549. [CrossRef]
14. Li, J. Machine learning-based analysis of defensive strategies in basketball using player movement data. *Sci. Rep.* **2025**, *15*, 13887. [CrossRef]
15. Hu, W. The application of artificial intelligence and big data technology in basketball sports training. *ICST Trans. Scalable Inf. Syst.* **2023**, *10*, e2. [CrossRef]
16. García-Aliaga, A.; Marquina, M.; Coterón, J.; Rodríguez-González, A.; Luengo-Sanchez, S. In-game behaviour analysis of football players using machine learning techniques based on player statistics. *Int. J. Sports Sci. Coach.* **2020**, *16*, 148–157. [CrossRef]
17. Román-Gallego, J.Á.; Pérez-Delgado, M.; Cofiño-Gavito, F.J.; Conde, M.Á.; Rodríguez-Rodrigo, R. Analysis and parameterization of sports performance: A case study of soccer. *Appl. Sci.* **2023**, *13*, 12767. [CrossRef]
18. Fernández, J. From training to match performance: An exploratory and predictive analysis on F. C. Barcelona GPS data. In Proceedings of the IEEE 16th International Conference on Data Mining Workshops (ICDMW), Barcelona, Spain, 12–15 December 2016.

19. Yunus, M.; Aditya, R.S.; Wahyudi, N.T.; Razeeni, D.M.; Almutairi, R.I. Talent Scouting and standardizing fitness data in football club: Systematic review. *Retos* **2024**, *62*, 1382–1389. [CrossRef]
20. Gu, C.; Varuna, D.S. Deep generative multi-agent imitation model as a computational benchmark for evaluating human performance in complex interactive tasks: A case study in football. *arXiv* **2023**, arXiv:2303.13323. [CrossRef]
21. Collins, G.S.; Moons, K.G.; TRIPOD Group. TRIPOD+AI: Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis, updated for artificial intelligence and machine learning. 2025, *manuscript in preparation*.
22. Hong, Q.N.; Fàbregues, S.; Bartlett, G.; Boardman, F.; Cargo, M.; Dagenais, P.; Pluye, P. The Mixed Methods Appraisal Tool (MMAT) version 2018 for information professionals and researchers. *Educ. Inf.* **2018**, *34*, 285–291. [CrossRef]
23. Shea, B.J.; Reeves, B.C.; Wells, G.; Thuku, M.; Hamel, C.; Moran, J.; Henry, D.A. AMSTAR 2: A critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ* **2017**, *358*, j4008. [CrossRef]
24. Leddy, C.; Bolger, R.; Byrne, P.J.; Kinsella, S.; Zambrano, L. The application of Machine and Deep Learning for technique and skill analysis in swing and team sport-specific movement: A systematic review. *Int. J. Comput. Sci. Sport* **2024**, *23*, 110–145. [CrossRef]
25. Val Vec, S.; Tomažič, S.; Kos, A.; Umek, A. Trends in real-time artificial intelligence methods in sports: A systematic review. *J. Big Data.* **2024**, *11*, 148. [CrossRef]
26. Sampaio, T.; Oliveira, J.P.; Marinho, D.A.; Neiva, H.P.; Morais, J.E. Applications of Machine Learning to Optimize Tennis Performance: A Systematic Review. *Appl. Sci.* **2024**, *14*, 5517. [CrossRef]
27. Barker, T.H.; Habibi, N.; Aromataris, E.; Stone, J.C.; Leonardi-Bee, J.; Sears, K.; Munn, Z. The revised JBI critical appraisal tool for the assessment of risk of bias for quasi-experimental studies. *JBI Evid. Synth.* **2024**, *22*, 378–388. [CrossRef]
28. Wolff, R.F.; Moons, K.G.; Riley, R.D.; Whiting, P.F.; Westwood, M.; Collins, G.S.; Mallett, S.; PROBAST Group. PROBAST: A tool to assess the risk of bias and applicability of prediction model studies. *Ann. Intern. Med.* **2019**, *170*, 51–58. [CrossRef] [PubMed]
29. Moola, S.; Munn, Z.; Tufanaru, C.; Aromataris, E.; Sears, K.; Sfetcu, R.; Mu, P.F. Chapter 7: Systematic reviews of etiology and risk. In *JBI Manual for Evidence Synthesis*; JBI Global: North Adelaide, Australia, 2020. [CrossRef]
30. Biró, A.; Cuesta-Vargas, A.I.; Szilágyi, L. AI-Assisted Fatigue and Stamina Control for Performance Sports on IMU-Generated Multivariate Times Series Datasets. *Sensors* **2024**, *24*, 132. [CrossRef] [PubMed]
31. Campaniço, A.T.; Valente, A.; Serôdio, R.; Escalera, S. Data's hidden data: Qualitative revelations of sports efficiency analysis brought by neural network performance metrics. *Motricidade* **2018**, *14*, 94–102. [CrossRef]
32. Chen, M.; Szu, H.; Lin, H.Y.; Liu, Y.; Chan, H.; Wang, Y.; Zhao, Y.; Zhang, G.; Yao, J.D.; Li, W.J. Phase-based quantification of sports performance metrics using a smart IoT sensor. *IEEE Internet Things J.* **2023**, *10*, 15900–15911. [CrossRef]
33. Demosthenous, G.; Kyriakou, M.; Vassiliades, V. Deep reinforcement learning for improving competitive cycling performance. *Expert Syst. Appl.* **2022**, *203*, 117311. [CrossRef]
34. Nagovitsyn, R.; Valeeva, L.; Latypova, L. Artificial intelligence program for predicting wrestlers' sports performances. *Sports* **2023**, *11*, 196. [CrossRef]
35. Rodrigues, A.C.N.; Pereira, A.S.; Mendes, R.M.S.; Araújo, A.G.; Couceiro, M.S.; Figueiredo, A.J. Using Artificial Intelligence for Pattern Recognition in a Sports Context. *Sensors* **2020**, *20*, 3040. [CrossRef]
36. Yu, A.; Chung, S. Automatic identification and analysis of basketball plays: NBA on-ball screens. In Proceedings of the International Conference on Big Data, Cloud Computing, Data Science & Engineering 2019, Honolulu, HI, USA, 29–31 May 2019. [CrossRef]
37. Marquina, M.; Lozano, D.; García-Sánchez, C.; Sánchez-López, S.; de la Rubia, A. Development and Validation of an Observational Game Analysis Tool with Artificial Intelligence for Handball: Handball.ai. *Sensors* **2023**, *23*, 6714. [CrossRef]
38. Ramanayaka, D.H.; Parindya, H.S.; Rangodage, N.S.; Gamage, N.; Marasinghe, G.M.; Lokuliyana, S. Spinperform–A cricket spin bowling performance analysis model. In Proceedings of the International Conference on Awareness Science and Technology 2023, Taichung, Taiwan, 9–11 November 2023. [CrossRef]
39. European Commission. *Ethics Guidelines for Trustworthy AI. High-Level Expert Group on AI*; European Commission: Brussels, Belgium, 2019.
40. OECD. *OECD AI Principles*; OECD: Paris, France, 2019.
41. UNESCO. *Recommendation on the Ethics of Artificial Intelligence*; UNESCO: Paris, France, 2021.
42. IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. *Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems*; IEEE: New York, NY, USA, 2019.
43. Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ* **2021**, *372*, n71. [CrossRef]

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