Jia et al. BMC Sports Science, Medicine and Rehabilitation (2025) 17:249
https://doi.org/10.1186/s13102-025-01294-0

BMC Sports Science, Medicine and Rehabilitation

RESEARCH

Open Access

A narrative review of deep learning applications in sports performance analysis: current practices, challenges, and future directions

Yunke Jia[1], Norli Anida Abdullah[2*], Hafiz Eliza[3], Qingbo Lu[4], Deyou Si[5], Hengwei Guo[6] and Wenliang Wang[7]

Abstract

Background The integration of deep learning techniques into sports performance analysis has significantly advanced athlete monitoring, motion tracking, and predictive modelling. These advancements have significantly improved the ability to assess performance, optimize training strategies, and reduce injury risks. However, despite notable progress, challenges remain in standardizing methodologies, ensuring model reliability, and enhancing real-time application across various sports disciplines.

Methods We conducted a systematic literature search of Web of Science Core Collection (WOS), China National Knowledge Infrastructure (CNKI), and Association for Computing Machinery Digital Library (ACM DL) for relevant studies published from 2015 to 2024, with no language restrictions. Eligible studies were those that explicitly applied deep learning techniques (such as convolutional and recurrent neural networks) to sports performance analysis tasks (e.g., action recognition and classification, motion detection and tracking, injury prediction) and reported their methodology and performance metrics. Key data, including sport type, application domain, and model type, were extracted for narrative synthesis.

Results A total of 51 studies met the inclusion criteria, covering a broad range of individual and team sports. Deep learning techniques in sports performance analysis were chiefly employed for action recognition, object detection and multi-target tracking, target classification, and performance or injury prediction. CNNs were the most common models for visual recognition tasks, while RNNs (including LSTMs) were frequently used for temporal sequence data. Most studies reported improved performance outcomes with deep learning; however, we observed considerable variability in data quality, model validation approaches, and cross-sport generalizability.

Conclusions Deep learning has demonstrated transformative potential in optimizing sports performance analysis by providing automated, data-driven insights. Future research should prioritize integrating multi-modal data sources,

First authorship: Yunke Jia

*Correspondence:
Norli Anida Abdullah
norlie@um.edu.my

Full list of author information is available at the end of the article

refining real-time analytics, and improving the adaptability of deep learning techniques across different sports contexts to support more precise and data-driven performance assessments.

Keywords Action recognition and classification, Deep learning, Injury prediction, Motion detection and tracking, Sports performance analysis

Introduction

Sports performance analysis serves as a critical pathway for enhancing athletic performance and optimizing training strategies [1–5]. It typically involves the collection of multi-source data from on-site or laboratory settings, the extraction of quantitative indicators, and comprehensive analysis [6, 7]. Existing performance analysis methods are broadly categorised into technical analysis and tactical analysis, with the former focusing on movement details and execution quality, and the latter evaluating individual tactical behaviour and team coordination efficiency [7, 8]. Traditional approaches are primarily based on manual video observation and wearable sensor technologies, which can provide foundational data support, yet often fall short in terms of real-time responsiveness, analytical granularity, and cross-sport generalizability [9, 10].

With the rapid evolution of computer vision and deep neural network technologies, traditional manual feature engineering is being increasingly supplanted [8]. As an end-to-end nonlinear modelling approach, deep learning enables automatic extraction of multi-level features from large-scale raw data, thereby enhancing performance in tasks such as action recognition, trajectory tracking, and outcome prediction [8, 11]. This shift has positioned data-driven modelling as a central research paradigm, moving away from rule-based expert systems towards adaptive learning from historical match records, training sessions, and physiological data, thereby improving predictive capability and model generalizability [4]. In parallel, advancements in multimodal data fusion techniques have enabled the integration of heterogeneous data sources—such as visual, electromyographic, and physiological signals—into a unified feature space, contributing to more robust task performance and enhanced training feedback [12]. These strategies not only improve model accuracy under complex conditions but also provide the foundation for real-time feedback and personalised training protocols. For instance, hybrid architectures combining convolutional neural networks (CNNs) and recurrent neural networks (RNNs) support joint modelling of spatial and temporal dynamics, allowing extraction of key kinesiology features relevant to athletic performance [11].

Given the rapid advancement of deep learning in sports performance analysis and the persistent methodological and practical challenges, a comprehensive synthesis of existing research is both timely and necessary. This review aims to systematically examine the current applications, limitations, and emerging trends of deep learning in this domain. By analysing studies published between 2015 and 2024, it seeks to provide an updated overview of the technological landscape, identify existing gaps, and propose forward-looking directions for future research. It is noteworthy that existing reviews often focus on specific technologies or narrowly defined application contexts. For example, Seçkin et al. [6] concentrated on wearable technologies in sports; Torres-Ronda et al. [7] reviewed data applications of tracking systems in team sports; and Cossich et al. [8] explored advancements in artificial intelligence, virtual reality, augmented reality, and data visualisation techniques for performance analysis. Additionally, Arzehgar et al. [9] conducted a scoping review on sensor-based motion analysis for injury assessment, while Rahimian and Toka [13] focused on optical tracking methods for players and balls in team ball sports. In contrast, this review begins with deep-learning techniques, systematically summarises their applications across multiple sports and task types, integrates disparate scenarios and methodologies overlooked by earlier work, and therefore offers the most recent and holistic perspective on the field.

Thus, the rest of the paper follows: Sect. "Methods" describes the review methodology, including database selection and screening criteria. Section "Results" summarizes key findings on publication trends, research hotspots, and the scope of included studies. Section "Discussion" discusses current limitations and future directions. Section "Conclusion" concludes with reflections on the potential, challenges, and opportunities for broader application of deep learning in sports contexts.

Methods

Literature search strategy

This review was conducted in accordance with published guidelines for narrative literature reviews [14, 15] and the PRISMA-Search guidelines [16–18] to ensure transparency and reproducibility. A systematic search was carried out across three electronic databases: Web of Science Core Collection (WOS), China National Knowledge Infrastructure (CNKI), and Association for Computing Machinery Digital Library (ACM DL). The search strategy was developed around the core concepts of "deep learning" or "neural networks" combined with "sports performance" or "sports analysis," using the Boolean operator AND. Additional keywords such as "recognition," "tracking," "classification," "detection," and "prediction" were incorporated to refine the query. The search was limited to publications from January 2015 to December 2024. No language restrictions were applied, although the majority of retrieved literature was in English and Chinese.

Inclusion and exclusion criteria

Literature screening was performed in two phases based on predefined eligibility criteria. Studies were included if they met the following criteria: (1) they clearly described the use of deep-learning methods for sports-performance analysis, including—but not limited to—action recognition, motion tracking, injury prediction or tactical decision-making; (2) algorithmic frameworks and performance evaluation metrics were explicitly reported; (3) the publication was a peer-reviewed journal or conference article; and (4) the study involved a systematic review of deep learning and sports performance analysis theories.

Exclusion criteria included: (1) studies that employed only conventional machine learning without deep learning; (2) absence of technical implementation details or performance metrics; (3) lack of relevance to sports performance or disconnection from specific application scenarios; and (4) abstracts, commentaries, retracted papers, and non-academic sources. All records were independently screened by four reviewers, with disagreements resolved through consultation with a fifth reviewer.

To further ensure the credibility of the evidence base, implicit quality thresholds were applied during the screening process. Included studies were required to demonstrate: (1) methodological transparency, with clear disclosure of data sources, sample sizes, and detailed model specifications; (2) result verifiability, through reporting of standardised performance metrics and comparison against established benchmarks; and (3) task relevance, by focusing specifically on competitive sports performance rather than general health or physiological outcomes.

Although no formal quality appraisal tool was employed, the application of rigorous inclusion criteria ensured that only studies with clearly reported technical implementation and validation were considered. This approach supports a structured and objective evaluation of current capabilities and emerging trends in deep learning–based sports performance analysis.

Synthesis and analysis

Following the screening process, a narrative synthesis was conducted for the included studies. Each study was systematically reviewed to document key elements such as sport type, task category, deep learning architecture, and core performance metrics. Based on these characteristics, studies were categorised into four thematic areas: action recognition, target detection and tracking, target classification, and performance prediction. In the Results and Discussion sections, we critically appraised the strengths and weaknesses of the included studies within each thematic category by focusing on data modalities and quality, model-architecture choices, and validation protocols and metrics. This approach preserved the flexibility of a narrative review while providing a transparent and structured evaluative perspective, aligning the methodological framework with the graded presentation of evidence reliability. It enables readers to accurately interpret the practical utility and remaining gaps in applying deep learning across diverse sporting contexts.

Results

A total of 6279 records were initially retrieved, including 6077 from WOS, 183 from CNKI, and 19 from the ACM DL. After removing 1 duplicate, 4 conference summary documents, and 77 retracted publications, 6197 records remained for the initial screening. Title and abstract screening led to the exclusion of 6119 records that were either irrelevant or did not meet the inclusion criteria, leaving 78 articles for full-text review. In the second screening phase, 27 articles were excluded due to insufficient methodological detail regarding deep learning or lack of relevance to sports performance analysis. Ultimately, 51 studies were included in the final review. The entire identification and screening process is illustrated in Fig. 1. These studies represent a diverse range of sports disciplines and deep learning application tasks, providing a comprehensive overview of current research trends in this field.

Overview of included studies

Distribution of research hotspots

To explore the key research themes and structural relationships in the application of deep learning to sports performance analysis, a co-word analysis of keywords was conducted using VOSviewer software based on the 51 included articles. A total of 24 high-frequency keywords (minimum occurrence ≥ 3) were identified. The keyword co-occurrence network is illustrated in Fig. 2. The analysis reveals that "deep learning" is the central term in this field, frequently associated with topics such as object detection, action recognition, and sports analysis. The clustering results indicate four primary research themes: image recognition, computer vision, neural network models, and sport-specific performance analysis. Overall, the research focus is evolving from algorithm-centric studies toward multi-task integration and application-oriented exploration.

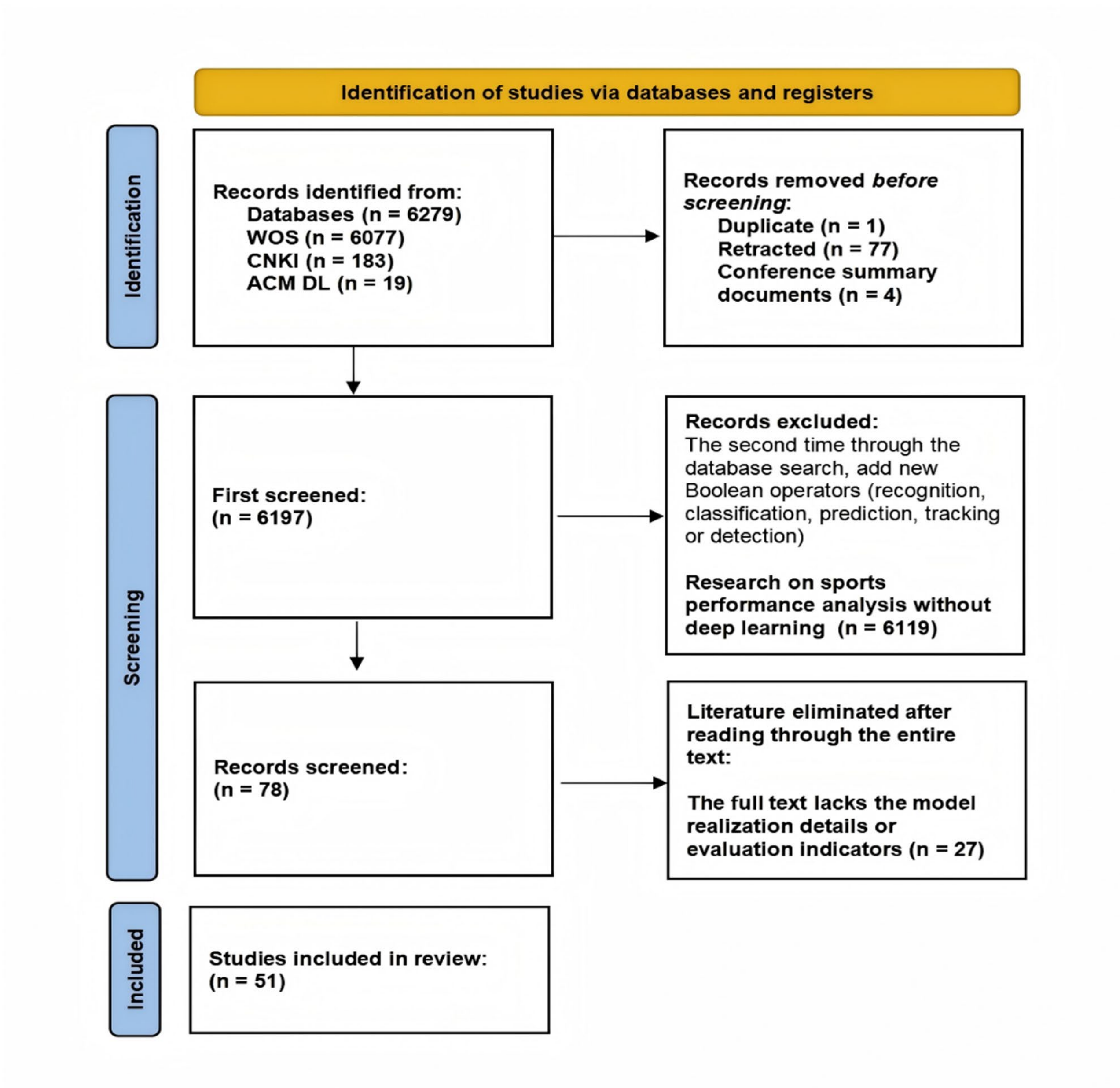

Fig. 1 PRISMA-S flow diagram for literature searches

Distribution by publication year

The publication years of the included articles span from 2015 to 2024, demonstrating a clear upward trend over the past decade. Only three studies were published between 2015 and 2017. From 2018 onward, publication volume increased steadily, with 6 articles from 2018 to 2019 and a notable rise thereafter. Between 2020 and 2023, 27 articles were published, reflecting growing academic interest in this topic. The surge corresponds with advancements in computer vision and sensing technologies. Notably, 15 studies have already been published in 2024, suggesting that the field continues to expand rapidly.

Sports types covered

The included studies span a diverse range of sports, covering both team-based and individual disciplines. Football and basketball emerged as the most frequently studied sports, with research primarily focusing on action recognition, classification, player or ball tracking, and match outcome prediction [13, 19–28]. Other sports, such as volleyball, cricket, and badminton also appeared in several studies. Volleyball-related research largely focused on technical action recognition and behaviour detection [29, 30], while cricket studies addressed action recognition and classification of traditional sports [31, 32], Badminton-related work emphasized racket stroke

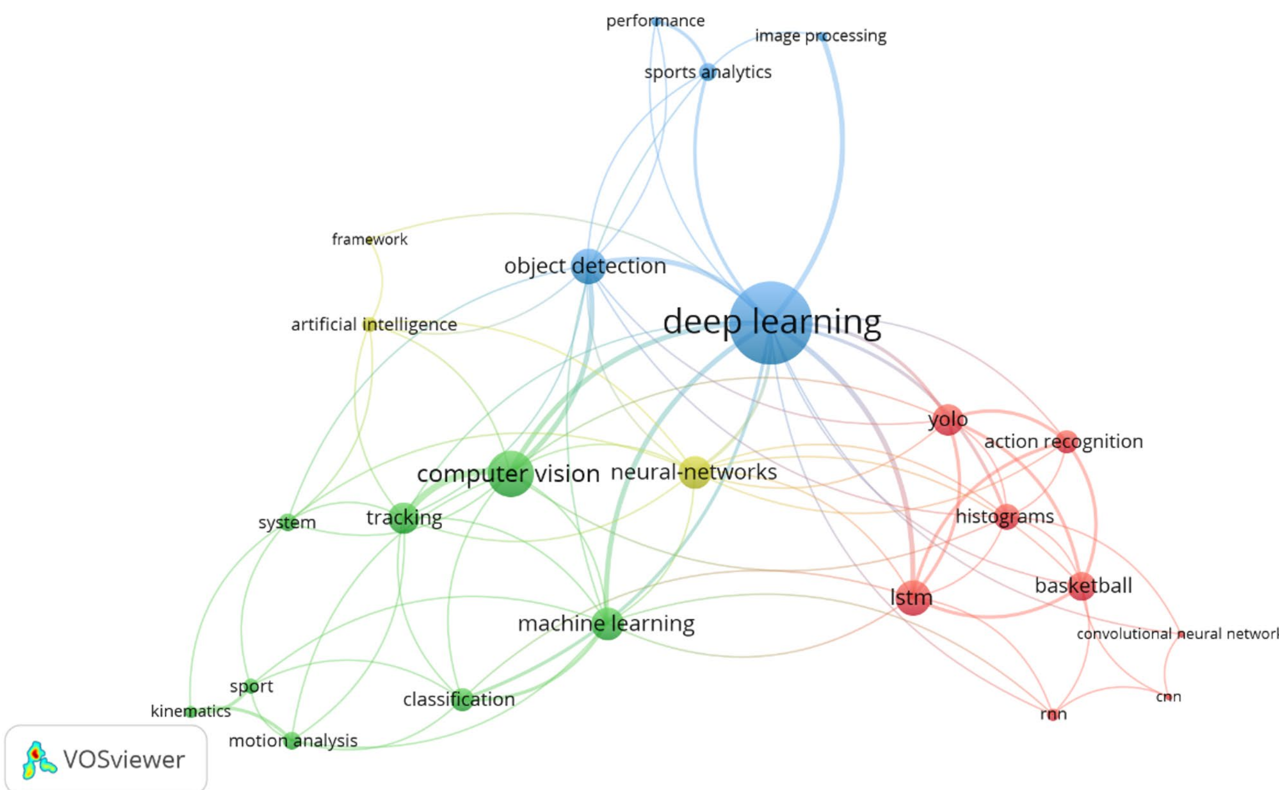

Fig. 2 Keyword co-occurrence network of deep learning in the field of sports performance analysis and research

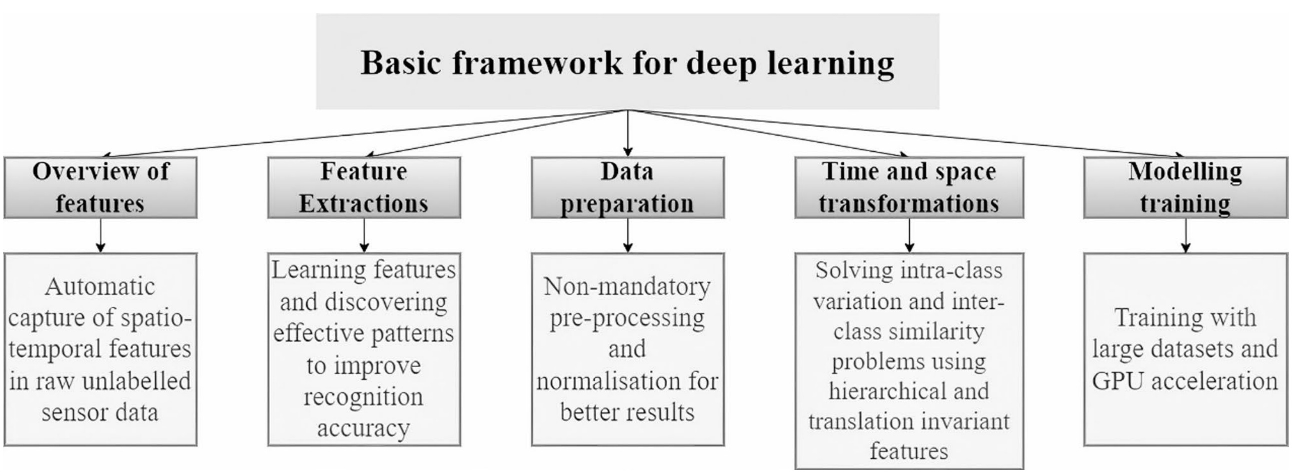

Fig. 3 Basic framework for deep learning

recognition and shuttlecock tracking [33–35]. A few studies also involved traditional physical activities [32] and multi-sport classification tasks [36, 37]. In summary, team sports dominate current research in this area, while individual sports, although less represented, demonstrate high applicability in specific recognition scenarios.

Overview of deep learning techniques

Definitions and features

Deep learning is a method within machine learning used for feature extraction, widely recognized for its effectiveness in pattern recognition and predictive modelling [38]. Unlike conventional approaches, deep learning autonomously extracts meaningful features from raw sensor data, reducing the need for manual rule definitions and expediting data processing (see Fig. 3). This capability has been instrumental in computer vision applications,

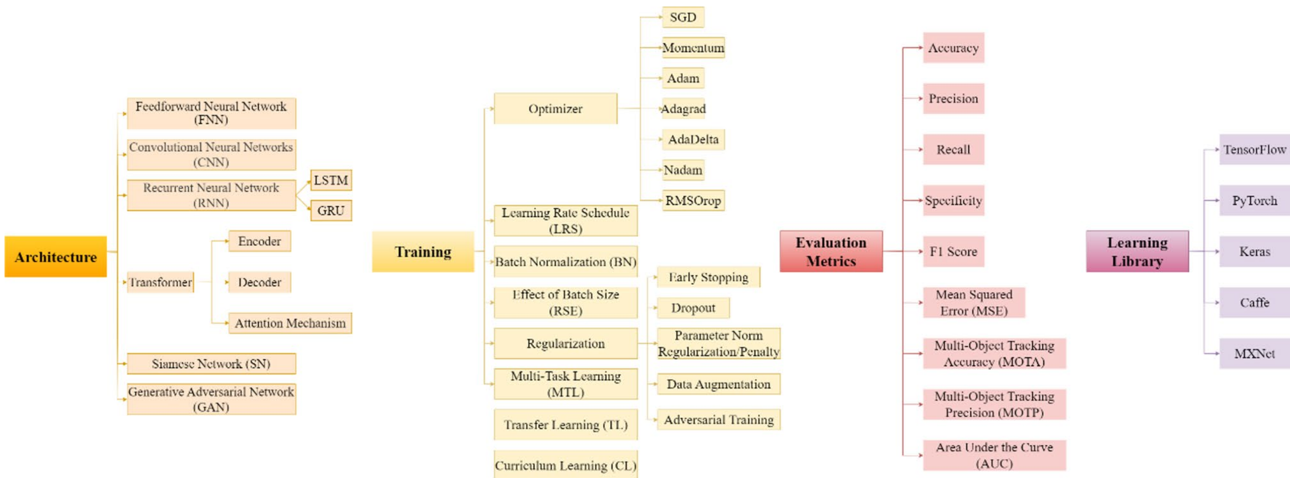

Fig. 4 Deep learning architecture diagram

Table 1 Advantages and disadvantages of common deep learning model architectures for sports performance analysis

Models	Advantages	Disadvantages
Convolutional Neural Network (CNN)	Efficient capture of spatial features for image processing tasks.	Difficult to handle long-term dependencies and sequential data.
3D Convolutional Neural Network (3D-CNN)	Extending convolution operations to three dimensions allows for more efficient feature capture.	High computational cost and memory demand.
Recurrent Neural Networks (RNN)	Able to process sequential data and capture temporal dependencies.	Suffers from the problem of vanishing gradients and difficulty in learning long-term dependencies.
Long Short-Term Memory Network (LSTM)	Solves the long-term dependency problem of RNN and can effectively handle long sequences.	High computational complexity and long training time.

particularly in sports performance analysis, where it processes raw images or video frames, classifies actions, detects movements, tracks motion, and predicts performance outcomes. By analysing the semantic structure of various sports activities, deep learning enables automated extraction of performance data for advanced analysis. The progress of computational power and large-scale data availability has significantly driven the adoption of deep learning as a standard approach in sports performance analysis [39].

Architecture

The architecture of deep learning systems typically encompasses structural design, training procedures, evaluation metrics, and tool parameters. Figure 4 provides an overview of commonly used architectures, training strategies, evaluation standards, and representative datasets employed for model training [40]. Among the various deep learning approaches applied in sports performance analysis, CNNs and RNNs are among the most frequently utilised techniques [41]. A comparative summary of prevalent deep learning architectures used in this domain is presented in Table 1 [42].

CNNs consist of interconnected layers designed to process structured image data, making them particularly effective for extracting information from static frames [43]. Since traditional 2D-CNNs are limited in capturing temporal information from video sequences, 3D-CNNs have been developed to analyse both spatial and temporal dynamics. Currently, CNNs are widely used in sports performance analysis for tasks such as image recognition, scene classification, and object detection [39]. RNNs, on the other hand, are well-suited for modelling sequential data, such as time series or raw sensor readings. These networks incorporate temporal layers to capture dependencies over time, making them particularly effective for action recognition. RNN models leverage recurrent connections to process complex variations in movement patterns, basing predictions on past observations and current inputs [44]. However, standard RNNs are susceptible to limitations in maintaining long-term dependencies, which constrains their applicability in extended activity sequences and temporal modelling [42]. To address these limitations, Long Short-Term Memory (LSTM) networks have been introduced. As a variant of RNNs, LSTMs incorporate memory cells that regulate the retention and forgetting of hidden states, effectively capturing long-range dependencies in sequential data [45]. By integrating memory cells and specialized gate mechanisms, LSTMs enhance the modelling of temporal structures, improving recognition accuracy [46]. Due to their effectiveness, LSTM models are increasingly employed in sports and computer vision research, including the analysis of athletic performance. Despite their advantages, LSTMs require extensive parameter tuning during training, contributing to increased computational complexity [47].

It is important to note that sports performance data often contain both spatial and temporal components, which pose challenges for any single deep learning model to effectively capture [48]. To address these limitations, many studies have adopted hybrid architectures. For instance, hybrid classifiers combining CNN and RNN have been applied to enhance the accuracy of next-action prediction [49]. Similarly, CNN-LSTM models have been developed to integrate spatial analysis with temporal modelling capabilities, enabling the precise identification of potential injury patterns and facilitating sports injury prediction [42]. These integrated frameworks help overcome the limitations of individual models and enhance robustness and adaptability to the complex nature of sports data.

Application of deep learning techniques in motion performance analysis

Athlete performance directly influences match outcomes. Deep learning techniques, particularly computer vision and optical sensor-based models provide actionable insights for sports teams, offering a deeper understanding of individual and team dynamics [50]. Analysing World Cup datasets—including player positions, goals, and fouls—illustrates performance variations across positions and playing styles, underscoring the value of deep learning in sports analytics. Deep learning technologies are integral to evaluating player performance across entire matches or seasons, encompassing positioning, movement patterns, and decision-making. Applications include activity recognition, target detection and tracking, classification, and prediction. These methods facilitate athlete positioning, identification, event classification, trajectory prediction, and match outcome forecasting, often visualized through annotated videos. Common performance evaluation metrics for deep learning models in sports performance analysis include (see Table 2): Accuracy (ACC), Positive Predictive Value (PPV), Recall Rate/True Positive Rate (TPR), True Negative Rate (TNR), F1 Score, Mean Squared Error (MSE), Multiple Object Tracking Accuracy (MOTA), Multiple Object Tracking Precision (MOTP), and Receiver Operating Characteristic Curve (ROC) and Area Under the ROC Curve (AUC) [51, 52]. Robust models enhance performance assessments, aiding coaches and researchers in identifying strengths and weaknesses, ultimately informing targeted training programs for athletic development [53].

Table 2 Deep learning-based key assessment metrics for sports performance analysis

	Function	Formula
ACC	The ratio of correct predictions to the total number of observations.	$ACC = \frac{\mathrm{TP+TN}}{\mathrm{TP+TN+FP+FN}}$
PPV	The ratio of true positive predictions to the total number of positive predictions.	$PPV = \frac{\mathrm{TP}}{\mathrm{TP+FP}}$
TPR	The ratio of true positive predictions to the total number of actual positive instances.	$TPR = \frac{\mathrm{TP}}{\mathrm{TP+FN}}$
TNR	The ratio of true negative predictions to the total number of actual negative instances.	$TNR = \frac{\mathrm{TN}}{\mathrm{TN+FP}}$
F1 Score	The harmonic means of precision and recall.	$F = 2 \times \frac{\mathrm{precision} \times \mathrm{recall}}{\mathrm{precision+recall}} = \frac{2\times \mathrm{TP}}{2\times \mathrm{TP+FN+FP}}$
MSE	The average of the squared differences between predicted and actual values.	$MSE\left(\mathrm{y}, \mathrm{y}'\right) = \frac{\sum_{\mathrm{i=1}}^{\mathrm{n}} (\mathrm{y_i} - \mathrm{y_i'})^2}{\mathrm{n}}$
MOTA	A metric that measures the accuracy of multi-object tracking algorithms.	$MOTA = 1 - \frac{\sum_t (\mathrm{FN_t + FP_t + IDSW_t})}{\sum_t \mathrm{GT}}$
MOTP	A metric that measures the precision of multi-object tracking algorithms in terms of object localization.	$MOTP = \frac{\sum_{t,i} d_{t,i}}{\sum_t c_t}$
ROC and AUC	A metric that measures the overall performance of a classification model across all possible classification thresholds.	

Note: When correctly classified, TP (True Positive) refers to correctly identified positive samples, and FP (False Positive) refers to falsely identified positive samples. When misclassified, FN (False Negative) refers to falsely identified negative samples, and TN (True Negative) refers to correctly identified negative samples. $\mathrm{y_i}$ represents the correct answer of the i-th data in a batch, and $\mathrm{y_i'}$ is the prediction given by the neural network. $t$is the time frame index, and GT refers to the ground truth object. FN (False Negative) is the number of ground truth objects that the method failed to detect. FP (False Positive) is the number of objects incorrectly detected by the method but not present in the ground truth. IDSW (ID Switch) occurs when an object reappears in view after being occluded and the network assigns it a new identifier, thus defining an ID switch. $c_t$indicates the number of matches in the t-th frame, and $d_{t,i}$indicates the Intersection-over-Union (IoU) between the target $i$in frame $t$and its designated ground truth object

Action recognition application of deep learning techniques in sports performance analysis

In sports performance analysis, the selection and standardisation of technical movements is the key to assessing training effectiveness. However, it is difficult to accurately judge details in high-speed movements with the naked eye, and manually labelling videos frame by frame is both time-consuming and laborious. Deep learning-driven real-time posture detection technology makes up for this deficiency: by automatically recognising the body postures of single or multiple athletes from visual data (extracting key joint positions such as elbows, knees, ankles, etc.), it is able to automatically capture the details of the technical movements, and assist coaches in evaluating and improving the quality of the movements. The core of the posture detection task is Human Action Recognition (HAR), which essentially decodes the 'when and what action' of an individual or a group from the data source, and is not only applicable to the detailed analysis of technical movements in single-player sports, such as badminton, tennis and skiing, but also widely used in football and basketball. HAR is not only applicable to the detailed analysis of technical movements in single player sports such as badminton, tennis, skiing, etc., but also widely used in the analysis of tactics in team sports such as football, basketball and volleyball [54]. For example, it can detect the specific movements performed by players in real time and their duration, and accurately calculate technical indicators. Compared to manual statistics, the HAR-based automated system can capture key information in training and competition scenarios more objectively and stably. Figure 5 illustrates the common model architectures and typical application scenarios of deep learning in the field of motion recognition. Currently, commonly used models include CNNs, RNNs, and LSTMs. Host et al. further provided a typical deep learning human motion recognition algorithm workflow (see Fig. 6), covering the complete technical process of training and testing data preparation, feature extraction, model training, data segmentation, result analysis, and performance evaluation [55].

Recent studies have reported numerous achievements in deep-learning-based action recognition for sport. In weightlifting, CNNs have been used to identify key postures during the lift, achieving accuracies above 95% [54]. In football, an augmented recurrent neural network was

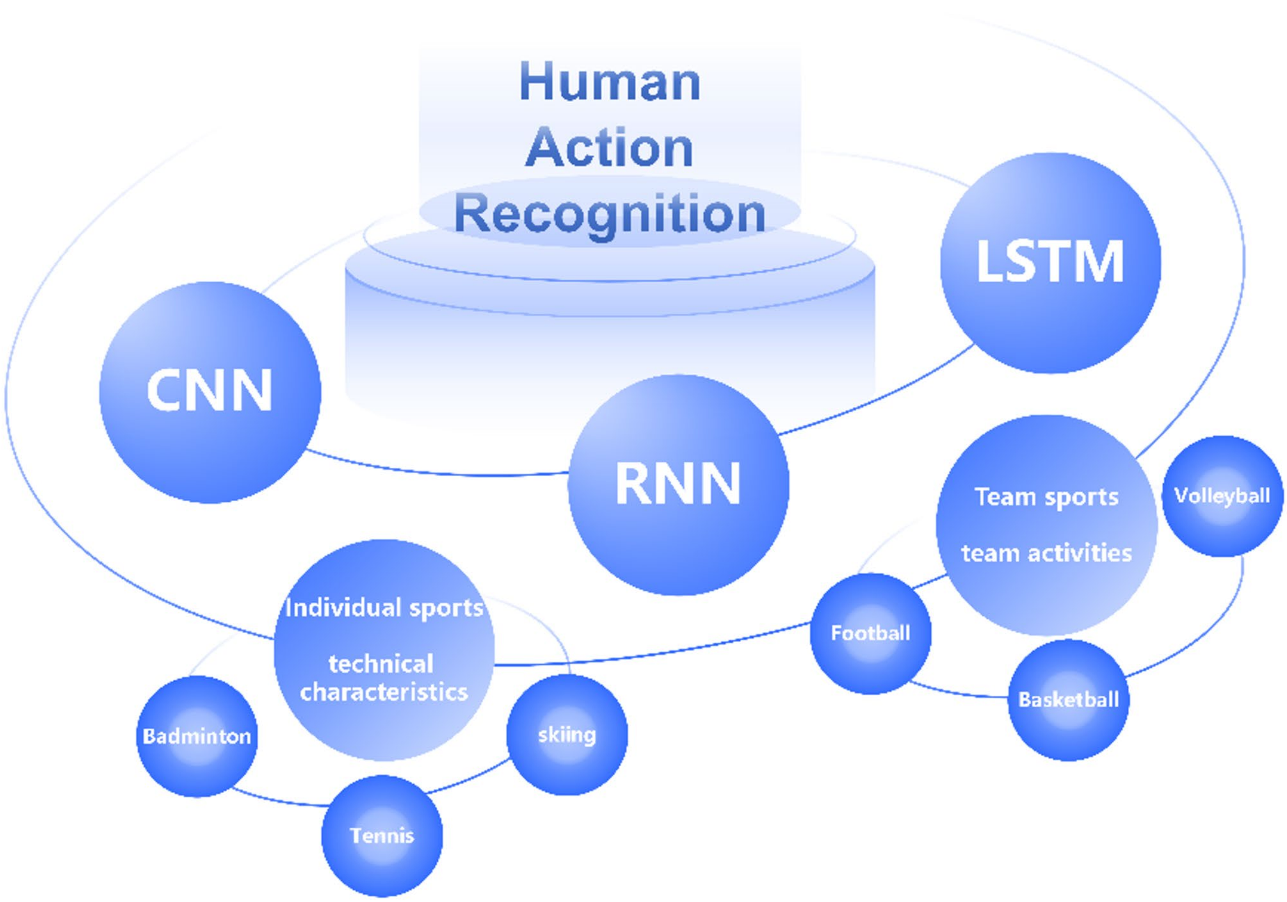

Fig. 5 Deep-learning architectures and application domains for sports action recognition

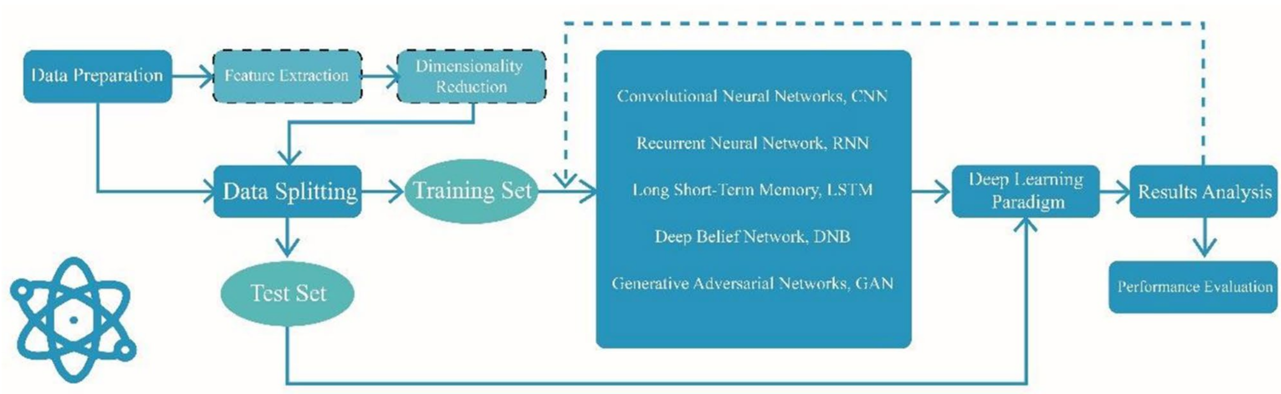

Fig. 6 End-to-end workflow for deep-learning-driven human motion recognition. (adapted from Host et al., 2022) [55]

Table 3 Comparative performance of deep-learning models for action recognition across sports

Sport	Sensor Modality	Data Representation	Model Architecture	Core Performance*	Main Strengths	Key Limitations
Weightlifting [54]	RGB sequence → FCN foreground segmentation (single athlete)	ROI frames 64 × 128, ≈ 612 frames / 4 classes	FCN + CNN	ACC 95.23–98.91%	Good noise suppression; high accuracy even with small samples; lightweight deployment	No temporal modelling; single viewpoint
Football [56]	RGB sequence (UCSD/CMU), single target, frame differencing	Continuous frames 224 × 224, 50–200 frames/segment	Inception-CNN + 2-layer GRU	ACC 95.4–96.3%	Joint spatial–temporal learning; end-to-end	Class imbalance; single camera; GRU serial processing
Multi-sport [57]	RGB sequence + skeleton (multi-target)	Video segments 50–200 frames + joint sequences	VGG + TS-Memory + Skeleton-DSN	ACC 98.5%	Long-term dependency capture; skeleton augmentation	Requires accurate pose estimation; real-time performance unverified
Badminton [34]	Stereo RGB 25 fps, dual view	Cropped ROI sequences (up to 1,775 frames)	Faster R-CNN + stereo localisation	mAP 0.982; ACC 97.2%	Non-wearable, high-precision 3-D footwork	Camera calibration required; single-player half court only
Basketball [22]	RGB (SpaceJam), YOLO cropping	15 ROI frames/segment	YOLOv4 + LSTM + T2 Fuzzy	ACC 96.45%	Illumination-robust; few frames needed	Detection dependent; heavy model
Volleyball [30]	RGB classroom 25 fps, GMM cropping	ROI frames 112 × 112 sequence	Improved C3D (dual resolution)	ACC after RGB + Crop fusion = 0.88	High accuracy and lightweight	Single-player scenario; requires segmentation
Basketball [23]	RGB 25 fps, multi-target, YOLO cropping	Sparse-sample clip stack	Enhanced YOLO + T2F-LSTM	ACC 99.3%	Multi-feature fusion, real-time	Large model; scene-specific
Cricket [31]	RGB 30 fps, single target	TD 10-frame clips, 100 × 100 px	TD-CNN + LSTM	Mean ACC 92.65%	Public dataset; end-to-end	Overall < 95%; few action types
Badminton [35]	RGB 25 fps, self-built 3 k images	60 frames/segment + 13 key points	YOLO-HG-Net + LightGBM	mAP50 96.1%	Lightweight, real-time; multi-scale	Small dataset; indoor only
Multi-sport [58]	RGB 1080p 25 fps, IoT stream	COCO/MPII single frames; batch inference	EfficientHRNet → DESNet	PCKh-MPII 90.9%	Lightweight, real-time; multi-scenario	Whole-body PCKh 90.9%; GPU required

applied to perform recognition of running patterns and tactical movements, again achieving high accuracy (95%) [56]. Ren et al. (2023) further improved pose recognition by incorporating skeletal-joint information, raising the mean accuracy for key-posture detection to 98.5% [57]. Caution is warranted when comparing performance scores across studies, as differences in datasets and evaluation protocols limit direct comparison. The 98% accuracy reported by one model may only be applicable to a specific test set and cannot be directly compared to the 95% accuracy reported by another study. At the same time, some high-performance metrics may stem from limited data sizes or insufficiently rigorous evaluation methods, with a potential risk of overfitting.

A comparison of the data modalities (Table 3) highlights the strengths and weaknesses of the different models. Vision-based methods capture rich contextual cues, often with high accuracy, but are more sensitive to occlusion and light changes [22, 23]. Skeletal keypoint approaches emphasise the common structure of movements and are computationally efficient, yet much depends on pose-estimation accuracy [22, 54, 57]. Sensor-fusion models excel at detecting subtle, occluded actions but limited insight into the wider tactical context [23, 34, 56]. To overcome the inherent limitations of single-modality inputs, recent studies have shifted towards multimodal fusion and coordinated architectural optimisation, thereby improving recognition robustness in complex settings [23, 30, 58]. Although deep learning has enhanced the automation and fine-grained resolution of sports-action recognition, real-world competition environments (characterised by frequent athlete occlusions, lighting variations and subtle interaction differences) still pose a challenge to model stability. To improve robustness under adverse conditions, recent studies have combined object-detection frameworks such as You Only Look Once (YOLO) with temporal models, achieving more reliable recognition in sports like basketball and badminton [22, 23, 35]. In addition, lifting 2-D video data to 3-D representations and effectively fusing heterogeneous data sources are emerging strategies for enhancing model generalisation, thereby strengthening deep learning systems' capacity to interpret complex sporting contexts and support decision-making.

Deep learning techniques for target detection and tracking in sports performance analysis

Target detection and tracking is a fundamental task in motion performance analysis, aiming to extract and monitor moving objects from continuous video sequences. The task is usually divided into two steps: spatio-temporal feature extraction and instance classification. CNNs are responsible for capturing appearance features, and classifiers such as Softmax generate bounding boxes for initial localisation [59]. Subsequently, target tracking trajectory data generated by cross-frame positional correlation lays the foundation for tactical and performance evaluation [13], and multi-object tracking (MOT) further improves the tracking accuracy [54].

Figure 7 visually reveals the three mainstream deep learning technology vectors in a circular layout. The first category is CNN detection methods represented by VGG-M, ResNet, etc., which focus on appearance information mining and have been widely used in fixed viewpoints or indoor environments to achieve high accuracy and stability [25]. For example, CNNs in football scenarios can quickly identify players and footballs to meet real-time analysis needs [21]; CNNs are used in basketball to accurately capture shooting trajectories and effectively assist in technical and tactical decisions [25]. Although CNNs perform well in spatial localisation and event detection in ball sports, their model generalisation ability is still limited to specific sports and specific scenarios. When CNN + IoU is used to track the position of the ball in a football match, the model may suffer from recognition confusion and misidentify the ball with the player if the colours of the uniforms of both teams are too similar [21]. The second class of methods is represented by the YOLO series of models (e.g., YOLOv3, YOLOBR, etc.), which achieve high frame rate and high recall with the help of a single-stage regression architecture, and are widely used in real-time referee assistance and event broadcasting systems [28]. A typical concrete process includes (see Fig. 8): YOLO detects targets in the first frame and generates bounding boxes, assigning unique trackIDs to each target. in subsequent frames, trackIDs are inherited through similarity matching of bounding boxes and combined with semantic information to

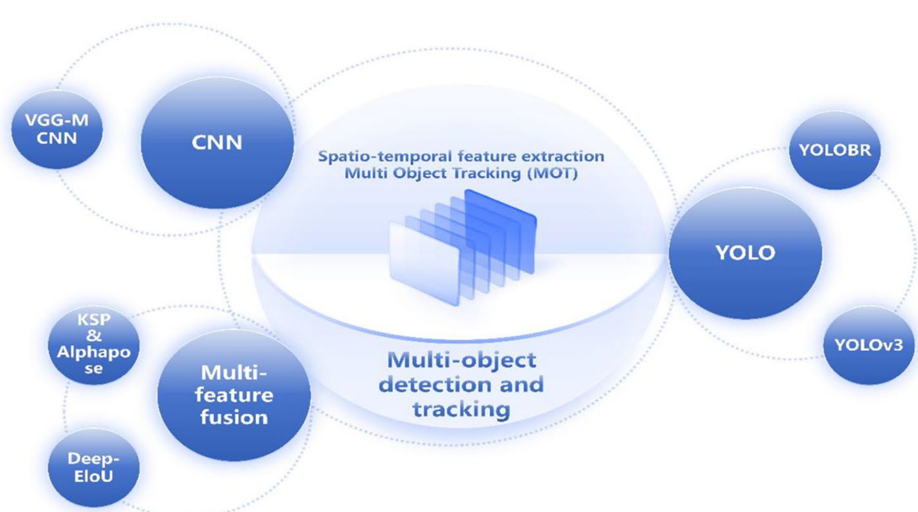

Fig. 7 Deep-learning paradigms for object detection and multi-target tracking in sport video

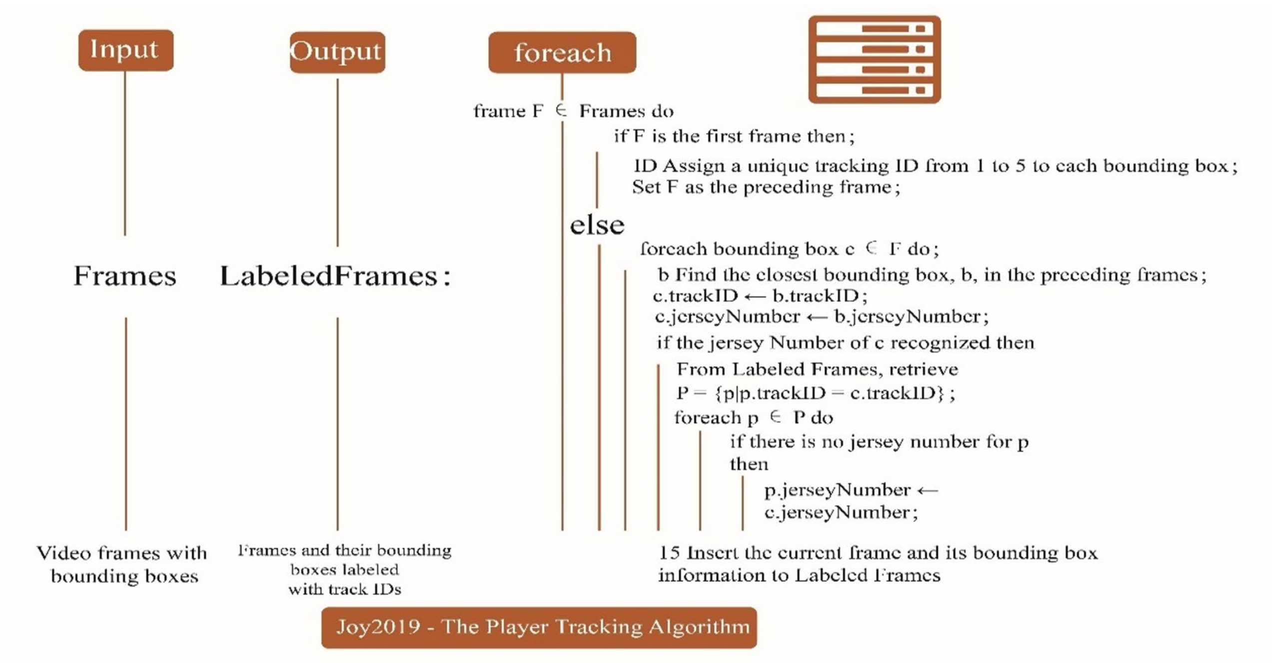

Fig. 8 YOLO-centric detection-to-tracking workflow for real-time basketball analysis. (adapted from Yoon et al., 2019) [28]

correct possible mismatches, thus achieving stable target tracking [28]. the core advantage of the YOLO model is its convolutional feature sharing and single-stage regression structure, which maintains high speed output and can be seamlessly integrated with lightweight data association strategies [28]. In addition, by optimising the loss function, the improved YOLO model also shows greater adaptability in dynamic environments. For example, the optimised YOLO in badminton can effectively cope with camera view angle changes and trajectory loss problems, and outperforms traditional methods in terms of accuracy and detection speed [33]. However, the YOLO family of methods still suffers from the trade-off between computational cost and detection accuracy, especially in complex scenarios with fast motion, multi-target interactions and frequent occlusions, and the identity drift of historical trajectory associations is still obvious [53]. The third category is the multi-feature fusion approach, which improves the robustness of MOT by integrating the information of appearance, motion trajectory and pose key points [60]. For example, Huang et al. (2024) integrated AlphaPose-based pose estimation with the K-Shortest Path (KSP) algorithm, markedly enhancing target continuity and tracking accuracy, as reflected by MOTA and Generalised Multi-Object Tracking Accuracy (GMOTA) under occlusion [37]. Thulasya Naik et al. (2024) used Euclidean Intersection over Union (EIoU) distance loss combined with jersey colour recognition to distinguish visually similar players and referees, which significantly reduces the risk of identity loss and mis-association in complex interaction scenarios [27]. Nevertheless, problems such as ambient lighting variations, self-occlusion and cross-camera coordinate alignment continue to challenge the practical applicability of existing multi-feature fusion methods [37]. As shown in Table 4, deep learning methods have successfully demonstrated their high performance and practicality in football, basketball and badminton; where CNNs combined with techniques such as IoU or EIoU have shown advantages in occlusion recovery and identity preservation [21, 27]. The YOLO architecture, on the other hand, is more suitable for real-time detection of small targets [33, 37].

Deep learning for target classification in sports performance analysis

Target classification in sports performance analysis refers to the accurate identification and labelling of key elements (e.g., athletes, balls, and regions associated with the event) in a video sequence [54]. The core of target classification lies in the accurate modelling and identification of Shape Dynamics, i.e., the morphological changes of targets in the spatio-temporal dimension [29]. As shown in Fig. 9, the current mainstream practice extracts spatial texture and temporal evolution features through the joint extraction of CNN and RNN (and its 3D variants); under this framework, improved networks such as VGG, Inception V3, Xception, DenseNet, etc. have already achieved significant results in a number of sports, such as cricket, traditional sports, badminton, tennis, and so on [32, 61, 62]. The local convolution mechanism of CNNs efficiently captures target contours and textures, while RNNs/3D-CNNs provide modelling

Table 4 Cross-sport evaluation of deep-learning approaches to object detection and tracking

Sport	Sensor Modality	Data Representation	Model Architecture	Core Performance	Main Strengths	Key Limitations
Basketball scoring [25]	Multi-camera HD video 25-30fps	5s/150-frame difference-fusion → Basketball Energy Image matrix	BEI + AlexNet P-HOG + SVM	Acc 94.59% F1 94.38% 45ms/frame	End-to-end real time; no explicit player/ball detection required	Low recall for "swish" shots; limited to scoring events
Football [21]	6 fixed cameras 1080p @ 25fps full-pitch stream	Sequential 107 × 107 patches	VGG-M CNN + IoU tracker	Acc 98.9% IoU 0.52 10 fps	Automatic ball localisation; recovers after occlusion	Loses ball when kit colours match; stands region ignored
Basketball [28]	Broadcast 720p @ 30 fps	16 k annotated boxes YOLO 13 × 13 tensor sequence	YOLOv3 + Joy2019 + NMS	Jersey-ID accuracy↑>100% 10 fps	Single-view detection and passing analysis	Pure 2-D; IDs confused on camera cuts
Badminton [33]	Stereo RGB 720p @ 24fps	18k frames → 416 × 416 grid	YOLOBR (pruned Tiny-YOLOv2)	Prec 96.3% Rec 94.6% 29.2 fps	4 MB weights; robust to small fast targets	Multi-lighting ghosting; 3-D reconstruction error
Football basketball handball ice hockey, etc. [13]	Multi-camera HD broadcast	107 × 107 patch sequence	Cascade-CNN/VGG/YOLO ,etc.	Accuracy 74–98%	Comparative survey; real-time potential	No unified benchmark; heterogeneous metrics
Basketball volleyball [60]	7 cameras 1080p + single view	4.8 k labelled frames K candidates/frame	YOLOv5 + Alpha-Pose + ReID + KSP	MOTA 0.83	Pose-trajectory fusion ↑ 6–12pp	High image-quality demand; costly setup
Basketball, football, volleyball [37]	Broadcast HD 25–30fps	1.6 M bboxes (training) online sequence	YOLOX-X + OSNet + Deep-EIoU	HOTA 77–85 14.6fps	Purely online; robust to large motions	Heavy computation; no 3-D output
Football [27]	ISSIA 6 cameras 1080p	MOG2 masks + bbox sequence	Unsupervised EIoU-Distance Loss	F1 0.91 6.6 fps	Training-free; automatic team separation	Illumination-sensitive; sub-real-time

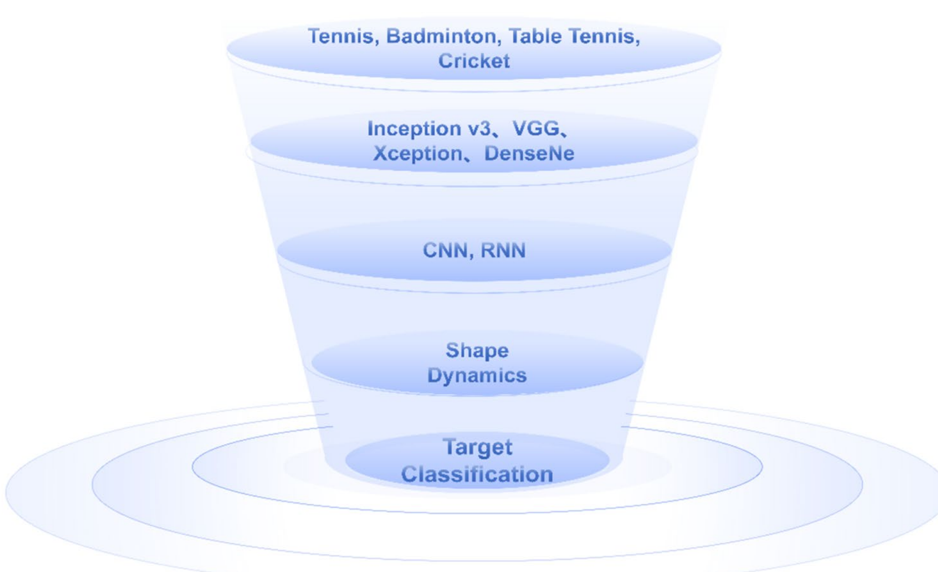

Fig. 9 Deep learning architectures and application areas for motion target classification

capabilities for morphing over time, which together fit the feature requirements of Shape Dynamics [63].

Advancements in deep learning have led to highly accurate classification models (see Table 5). For example, one study used CNN to classify football videos with an accuracy of 97.4% [24]; in the classification of batting/pitching scenes of cricket matches, the use of a pre-trained AlexNet model combined with data augmentation

Table 5 Performance comparison of deep learning models for target classification in sports

Sport	Sensor Modality	Data Representation	Model Architecture	Core Performance	Main Strengths	Key Limitations
Cricket [61]	YouTube, 30 fps, multi-camera	6800 × 227^2 RGB frames, 5 classes	AlexNet + TL-FC	Acc 99.26%	Very high accuracy; fast convergence with small samples	Only 5 classes; single visual modality
Traditional Bengali Sports [32]	YouTube HD, 30 fps, 5 s clips	TBSV: 500 clips × 20 frames 128^2 RGB, 5 classes	VGG19 + LSTM	F1 99%, Acc 92%	Transfer learning; few parameters; real-time capable	Small dataset; GPU dependent
Volleyball [29]	Broadcast match video, 30 fps	Volleyball: 3 segments 224^2 RGB + optical flow	CNN + HBR	Acc 78.3%, 85.46%, 89.89%	2D/3D spatiotemporal fusion; controllable compute load	Weak single-action accuracy; risk of overfitting
Ergonomic Posture [64]	Kinect RGB-D, 30 fps	9000 × 4 frames 60^2 skeleton & color	4-Route CNN + FC	F1 88.4%	Occlusion-robust; single-camera deployment	Small data; weak cross-scene generalisation
Football [24]	Broadcast RGB, 30 fps	SoccerNet-v2: single 224^2 RGB frame + 17-joint skeleton	CNN + GCN	Acc 97.4%	Appearance & skeleton complementarity; no wearables	Depends on skeleton estimation; generalisation unverified
Multi-sport (20 classes) [62]	Broadcast RGB, 30 fps (+ 480 p)	DeepSports-VDS: 80 frames 299^2 RGB	Dual-branch CNN + LSTM	Acc 98.0% (HD) / 88.0% (LD)	Robust to resolution change; leading accuracy	Heavy computation; heterogeneous samples

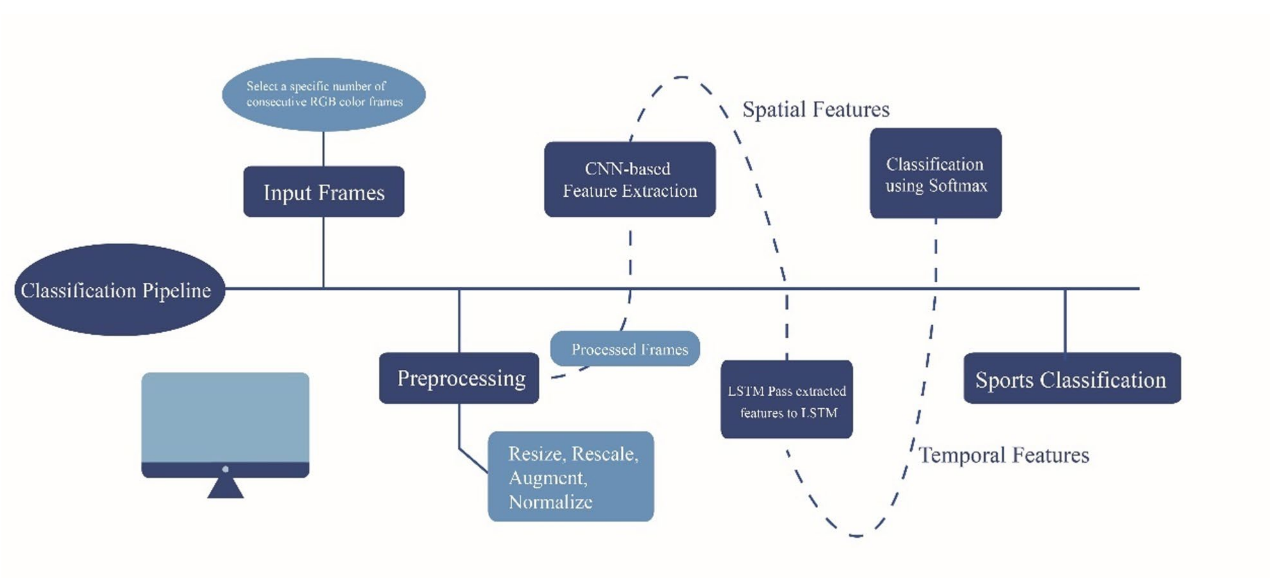

Fig. 10 End-to-end multi-branch CNN–LSTM pipeline for sports target classification. (adapted from Sarma et al., 2021) [32]

achieved a similarly high accuracy of 99.26% [61]; in contrast, for the more complex task of recognising volleyball technical movements, the model's accuracy on the HMDB51 dataset only achieved an accuracy of approximately 78.3% [29]. This difference in performance reflects that the complexity of different sports and tasks has a significant impact on model effectiveness. In situations where the visual pattern is relatively homogeneous and highly discriminative (e.g., discriminating batting and bowling actions in cricket), the depth model is more likely to achieve near-perfect performance [61]; whereas when the scene involves multiple athletes, a fast sphere, and frequent interactions (e.g., volleyball rounds), the difficulty of recognition increases significantly, and the model performance consequently decreases [29]. This indicates that task difficulty and category distribution directly determine the upper limit of model performance. However, the high score figures are not necessarily meaningful for side-by-side comparisons because the datasets, evaluation protocols, and positive-to-negative sample ratios used in each study are different, and a little slack may lead to overfitting. Therefore, when quoting performance metrics, it is important to report sample size, class baseline and cross-validation strategy to avoid the 'illusion of precision'.

This has prompted researchers to seek strategies such as transfer learning to mitigate these challenges. By utilising models pre-trained on large general-purpose image datasets such as ImageNet and applying their fine-tuning to the sports video domain, it is possible to reduce training overheads and improve performance in small-sample scenarios. Commonly used pre-trained models (e.g., Inception v3, VGG, Xception, and DenseNet, etc.) have demonstrated excellent image recognition capabilities on benchmarks such as ImageNet, and migrating them to sports classification tasks tends to achieve high accuracy [32]. Figure 10 illustrates the deep learning sports target classification process proposed by Sarma et al. (2021): video frame sequence → convolutional feature extraction → temporal modelling → output classification results [32]. This architecture represents a typical deep learning pipeline for sports video target classification and helps to illustrate how pre-trained models can be combined with specific tasks. However, there are significant differences between different sports, and even with transfer learning as a starting point, fine tuning for the target domain is still required. Recent research has attempted to introduce multi-frame and multi-path CNN architectures, which have improved recognition accuracy under challenging conditions [64]. Similarly, advanced neural network architectures combining keyframe selection, deblurring and multi-scale feature extraction have been shown to be effective in improving classification performance, especially in fast-paced and low-resolution sports videos [62].

Deep learning has driven the automation of sports target classification, but its true value comes from the three-dimensional balance of performance-resources-generalisation. When models can maintain > 90% accuracy in complex events, run in real-time on regular GPUs, and achieve consistent performance on new events with little incremental learning, sports analysts will be able to gain solid insights that can feed into training decisions and injury prevention.

Fig. 11 Deep learning architecture classification and its application areas in sports performance prediction

Prediction application of deep learning techniques in sports performance analysis

Predictive applications in sports performance analysis leverage deep learning techniques to forecast match outcomes, movement trajectories, and injury risks. By analysing historical data, these methods provide valuable insights that inform training plans and tactical decisions, enabling optimized performance strategies. Deep learning-based predictive frameworks typically incorporate specialized neural networks such as CNN, RNN, LSTM, and GRU, which are trained, validated, and tested on sports datasets to generate accurate predictions [26]. This approach improves the accuracy of decision making, facilitates data-driven training programmes, and is widely used in the areas of sports trajectory, future game performance, and injury risk (see Fig. 11).

Deep learning's predictive capabilities extend across key areas such as match outcomes, movement analysis, and tactical decision-making. Analysing team performance, spatial positioning, and movement patterns allows for forecasting future performance trends. Table 6 summarises the performance of the main models in sports prediction in different sports scenarios. In team sports, LSTM models effectively predict basketball actions, including shooting, passing, dribbling, and fouling, with accuracy rates of 80%, 71%, 83%, and 72%, respectively, further validating their effectiveness in movement trajectory analysis [19]. Yücebaş (2022) analysed data from the 2010 and 2014 FIFA World Cups using a multilayer artificial neural network and found that defensive characteristics of goalkeepers and defenders had the most significant impact on match outcomes. The model achieved an average accuracy of 81.34%, recall of 87%, and an F1 score of 0.84 [50]. EL-KASSABI et al. (2020) used a GRU model with 50 hidden states to predict the effects of environmental conditions on marathon performance, considering training metrics such as distance, duration, pace, and elevation. They found that training volume and intensity significantly influenced performance, while nutrition and hydration were critical for optimizing performance. Rest and recovery were essential for preventing overtraining and improving performance. Psychological factors, motivation, and genetic factors (e.g., fibre type and aerobic capacity) also affected athletic potential and environmental conditions such as temperature and humidity impacted performance. The model demonstrated the best performance with faster training loss reduction, a Mean Absolute Percentage Error (MAPE) of 0.052, and an accuracy of 95%. After

Table 6 Comparison of deep learning sport performance prediction methods in different sports

Sport	Sensor Modality	Data Representation	Model Architecture	Core Performance	Main Strengths	Key Limitations
Basketball (NBA) [19]	15–30 fps broadcast video	Spatio-temporal-attention key frames, 64 × 48 RGB	CNN + attention LSTM	Shooting 80%, Passing 71%, Dribble 83%, Foul 72%	Contact-free, real-time analytics; mean accuracy ≈ 77%	Low Pass/Foul precision; limited samples; high computational load
Marathon (Boston 2017) [65]	GPS + heart-rate logs	≤ 5-month variable-length training sequences	GRU-HS50	MAPE 0.052; accuracy 95%	Captures long-term load; no costly sensors	Single-race dataset; overfitting after 55 epochs
Football (1872–2018 & WC-2018) [26]	Match-event logs	≈ 40 events × 10 features per match	2-layer LSTM + Softmax	AUC 0.74; group-stage hit rate 63%	Open data; temporal model outperforms logistic regression	Knock-out accuracy drops; feature-sensitive
Football (WC-2010 & 2014) [50]	Player statistics (26 metrics)	≈ 1 600 vectors, grouped by position	3–4-layer fully connected DNN	Precision 81.34%, Recall 87%, F1 0.84; AUC 0.798	Reveals individual impact on result; low cost	Only two tournaments; few knock-out samples
Football (J1 League, 31 matches) [20]	10 fps monocular video + audio	60-frame RGB + spectrogram; 20-node graph	2-layer GCN → GRU-based GCRNN + BNN	AP 0.948; F1 0.858; mTTE 3.8 s	Low-cost sensing; predictive uncertainty output	Only 400 clips; limited xG transferability
Collegiate multi-sport [66]	Lab tests + questionnaire	> 20-dim strength/flexibility/injury-history vectors	Random Forest	Injury-risk accuracy 79%	Automatically finds risk factors; handles nonlinearity	Time-consuming tests; student-only sample
Badminton (singles) [67]	24 fps match video	48-frame RGB clips + court masks	UNet + YOLOv5 + TSM-ResNet50	Position 0.991; Action 0.902; Gesture 0.950	Single-camera, triple real-time tasks; low compute	Only 20 matches; generalization unverified

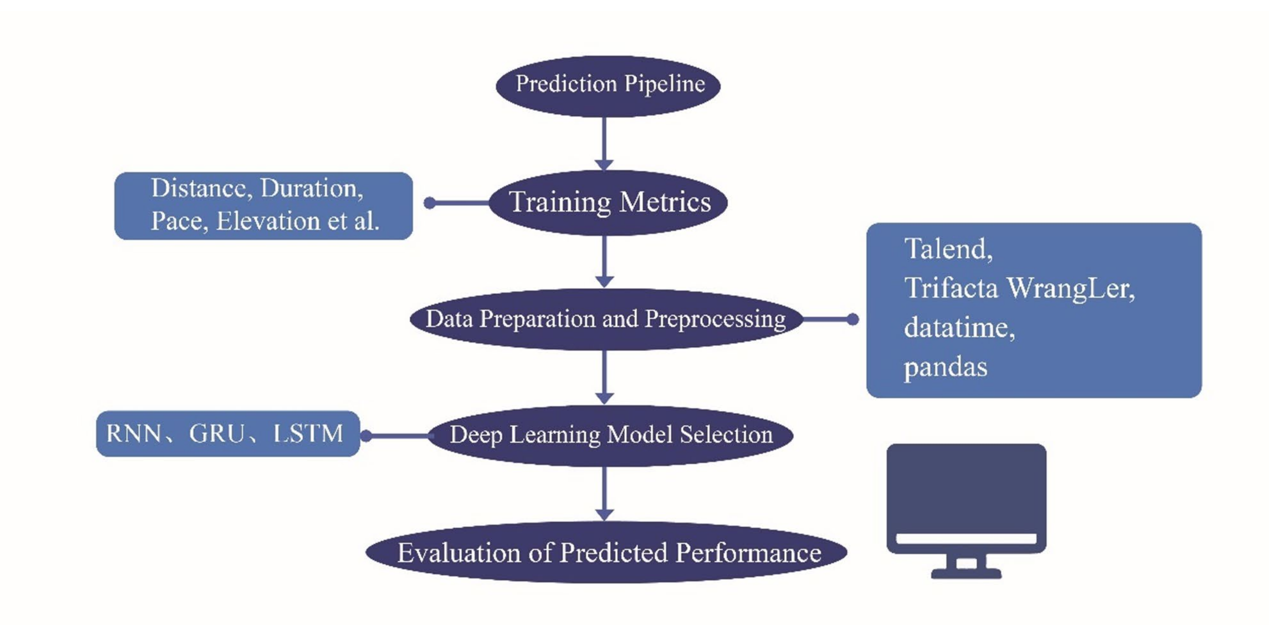

Fig. 12 Deep learning driven end-to-end workflow for sports performance prediction. (adapted from EL-KASSABI et al., 2020) [65]

removing outliers, the accuracy of RNN and LSTM models was 94% and 89%, respectively (For the flow chart, see Fig. 12) [65]. Beyond performance forecasting, deep learning contributes significantly to athlete health management. By examining biomechanical data, physiological indicators, and training loads, AI-driven models can assess injury risks and provide proactive health strategies. Real-time biomechanical monitoring enhances personalized injury prevention strategies, allowing athletes to mitigate risk factors through targeted interventions [66]. Recent advancements have further refined predictive applications in sports analytics. Football shooting event prediction has benefited from graph-based models integrating player relationships with audio-visual data, achieving high precision and outperforming traditional methods [20]. The introduction of Bayesian Neural Networks (BNN) for probabilistic prediction has significantly improved precision rates, reaching 0.948 with an F1 score of 0.858, demonstrating the potential for deep learning in strategic decision-making. The integration of Large Language Models (LLM) with computer vision has also expanded predictive analysis capabilities. Advanced frameworks now enable intelligent reasoning and real-time match analysis, significantly enhancing athlete and coach decision-making processes. For example, the application of deep learning in badminton match analysis has yielded accuracy rates of 0.991 for position recognition, 0.902 for action recognition, and 0.950 for gesture recognition, illustrating the precision of AI-driven sports analysis [67].

It is important to emphasise that it is not sufficient to judge models based on a small number of high precision metrics. The models in Table 6 report metrics such as ACC, Precision, and F1, but these high values may hide data leakage or sample bias issues. For example, if the same athlete or game clip is present in both the training and test sets (data leakage), the model may incorrectly learn the 'test answer'; or if certain events (e.g., successful shots) are too common in the data, the model may achieve high accuracy even with a simple model. If the validation process is not standardised, e.g. no independent test set, lack of cross-validation, hyper-parameter tuning directly to the test set, then the metrics will be grossly over-represented. Therefore, in addition to reporting overall values such as ACC/F1, the robustness and sensitivity of the model needs to be assessed by means of confusion matrices, ROC curves, or leave-one-out methods to verify whether the model is truly capable of generalisation. The motion prediction task itself is highly complex, and it is difficult for existing deep learning frameworks to comprehensively cover all the influencing factors. Existing models are mainly based on spatio-temporal sequence features, which have limited ability to capture the non-linear dynamics of the game. The physiological and psychological states of athletes (e.g., fatigue, stress) are often not directly accessible through visual/sensor data, and tactical interactions involve collaborative decision-making with multiple players and balls, which are beyond the capability of a single sequence model. Although approaches such as multimodal data fusion, attentional mechanisms, and hybrid models have expanded the model horizon to some extent, they still struggle to accurately simulate the complex causal relationships in real games.

Discussion

Limitations of deep learning techniques for sports performance analysis

Progress has been made in the application of deep learning in sports performance analysis, but there are still many limitations. Firstly, current research focuses on popular sports such as football, basketball and tennis, with insufficient coverage of other sports scenarios. Second, deep learning models usually rely on a large amount of high-quality labelled data, which is difficult to obtain in cold projects; their high arithmetic demand also raises the implementation cost. In addition, there is a lack of unified open datasets and evaluation standards in this field, and different studies tend to adopt their own datasets and protocols, which makes it difficult to reproduce or compare the results horizontally. As well, the black-box nature of the models brings a lack of interpretability, making it difficult for coaches and analysts to understand and trust AI decisions. If the training data is biased in terms of gender, age or sport level, model predictions may be systematically inaccurate for certain groups. Therefore, there is a need to introduce interpretable AI methods (e.g., visualisation or local interpretation techniques) to improve transparency, and to mitigate the risk of bias by using fairness assessment metrics or more diversified data to prudently validate model reliability [2, 53, 68, 69].

Future research directions and trends

Although existing research has laid a solid foundation for the application of deep learning in sports performance analysis, a review of the literature reveals several unresolved issues and future trends (see Fig. 13). Future research trends can be prioritised according to achievability and impact. In the short term, the focus could be on optimising existing technologies.

- Target Tracking: Rapid player movements and frequent occlusions pose challenges in multi-person tracking. Traditional linear motion-based tracking struggles with appearance-based occlusions, leading to identification switches. Future research should focus on developing frameworks that incorporate

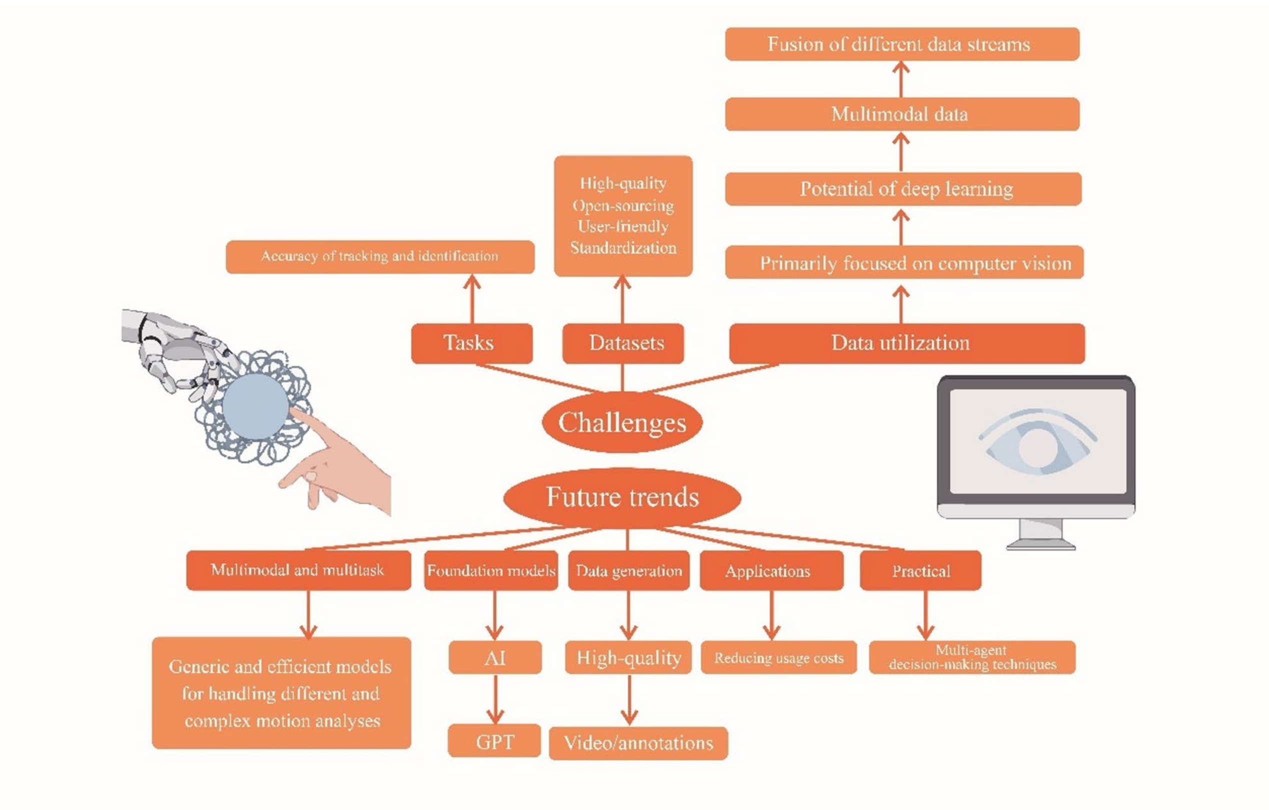

Fig. 13 Future trends of deep learning in sports performance analysis, from (ZHAO et al., 2025) [48]

identity-specific features to improve tracking accuracy [2, 53, 68, 69].

- Action Recognition: One of the most critical challenges in action recognition is accurately detecting athletes' actions when they exit and re-enter the frame, especially under varying lighting conditions, resolutions, and occlusions. Trajectory-based methods address some occlusion issues but are insufficient for ball-related tracking. Advanced AI models should be developed to predict ball trajectories with higher precision, leveraging extensive datasets for improved accuracy [2, 48, 53, 68, 69].
- Predictive Modelling: Forecasting match outcomes must consider a variety of influencing factors, including home and away conditions, player psychology, and environmental variables such as weather. Future research should explore more effective temporal segmentation techniques, improved word embedding strategies, and advanced models to capture psychological and physiological states, thereby enhancing predictive accuracy [2, 48, 53, 68, 69].
- User-Friendly AI Applications: The complexity of deep learning models often limits their accessibility to sports professionals without a technical background. Simplified tools and intuitive interfaces will be key to bridging this gap. The adoption of AI-powered analysis platforms such as Spiideo (https://www.spiideo.com/), iSportsAnalyse (https://www.isportsanalysis.com/rugby-video-analysis.php), and Gemmo AI (https://gemmo.ai/) will enable non-experts to perform in-depth performance evaluations. Similarly, sports professionals may increasingly leverage FIFA's Electronic Performance and Tracking Systems (EPTS) (https://inside.fifa.com/innovation) like TRACAB Gen5 (https://tracab.com/products/tracab-technologies/) and open-source frameworks available on platforms like GitHub (https://github.com/) to develop customized performance analysis tools.
- Future Paradigms in Deep Learning: In the long term, exploring novel learning paradigms will expand research horizons. Methods such as blended learning, combinatorial learning and simplified learning show potential in other domains, and researchers should evaluate their applicability in motion analysis. By critically introducing these cutting-edge methods, it is expected that model capabilities will be enhanced in terms of efficiency, accuracy, and interpretability [69].

Conclusion

Deep-learning techniques have proved highly valuable in sports-performance analysis, raising the accuracy of classification, action recognition, and tracking. Leveraging large-scale spatiotemporal datasets, they have improved action classification across diverse sports, while trajectory-based prediction models now inform tactical decisions and optimise team strategies. Most evidence, however, stems from controlled experiments, and model applicability and stability in real-match settings remain to be established. Although coaches can employ these tools to support training and match analysis, their output must be interpreted alongside domain expertise to avoid over-reliance. Because current models depend heavily on extensive annotated data, sustained collection of diverse, high-quality datasets—coupled with the integration of physiological and other multimodal inputs—is essential to enhance reliability. A further obstacle is the absence of unified algorithms and standards across sports: variability in data sources, equipment, and contexts hinders the development of generalisable models. Consequently, scalable deep-learning frameworks tailored to specific sporting scenarios, together with cross-disciplinary data standards and evaluation protocols, are urgently needed. Future research should prioritise algorithm interpretability, lightweight real-time deployment, and robust multimodal fusion. In practice, coaches are encouraged to participate actively in data collection and annotation, communicate closely with technical teams, and align model outputs with traditional insights; sports scientists should promote open data sharing and establish harmonised processing pipelines and evaluation criteria; and AI developers should focus on cross-scenario generalisability, explainability, robustness, and real-time performance. Only through coordinated efforts among coaches, researchers, and developers can deep learning be fully integrated into sports analytics, delivering precise, data-driven support for training and competitive decision making.

Abbreviations

WOS	Web of Science Core Collection
CNKI	China National Knowledge Infrastructure
ACM DL	Association for Computing Machinery Digital Library
CNNs	Convolutional Neural Network
RNNs	Recurrent Neural Network
DNN	Deep Neural Network
DRL	Deep Reinforcement Learning
LSTM	Long Short-Term Memory
3D-CNN	3D Convolutional Neural Network
ACC	Accuracy
PPV	Positive Predictive Value
TPR	True Positive Rate
TNR	True Negative Rate
MSE	Mean Squared Error
MOTA	Multiple Object Tracking Accuracy
MOTP	Multiple Object Tracking Precision
ROC	Receiver Operating Characteristic Curve
AUC	Area Under the ROC Curve
TP	True Positive
FP	False Positive
FN	False Negative
TN	True Negative
FP	False Positive
FN	False Negative
IDSW	ID Switch
YOLO	You Only Look Once
KSP	K-Shortest Path
GMOTA	Generalised Multi-Object Tracking Accuracy
EIoU	Euclidean Intersection over Union
IoU	Intersection-over-Union
HAR	Human Action Recognition
MOT	Multi-object tracking
MAPE	Mean Absolute Percentage Error
BNN	Bayesian Neural Network
LLM	Large Language Model
EPTS	Electronic Performance and Tracking System

Acknowledgements
None.

Author contributions
JIA conceived the research topic, developed keywords, collected data, synthesised findings, prepared Figs. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 and 13; Tables 1, 2, 3, 4, 5 and 6, wrote the manuscript, and supervised the project. NORLI and ELIZA provided research direction and ideas; SI, LU, GUO and WANG carried out systematic literature searches and critical appraisal of analyses and data collection. All authors reviewed the manuscript.

Funding
No research funding was received to support this work.

Data availability
All the data analysed in this study are from public literature and previously published materials. These articles have been cited in the text and references.

Declarations

Ethics approval and consent to participate
Not applicable to our study, being a narrative review article.

Consent for publication
Not applicable to our study, being a narrative review article.

Competing interests
The authors declare no competing interests.

Author details
[1]Institute for Advanced Studies, Universiti Malaya, Kuala Lumpur 50603, Malaysia
[2]Centre for Foundation Studies in Science, Universiti Malaya, Kuala Lumpur 50603, Malaysia
[3]Faculty of Sports and Exercise Science, Universiti Malaya, Kuala Lumpur 50603, Malaysia
[4]Police Officer Academy, Shandong University of Political Science and Law, Jinan 250014, China
[5]School of Mechanical and Automotive Engineering, Liaocheng University, Liaocheng 252059, China
[6]School of Mechatronic Engineering and Automation, Shanghai University, Shanghai 200444, China
[7]Department of Athletic Training, Shanxi Sports Vocational College, Taiyuan 030024, China

Received: 14 May 2025 / Accepted: 1 August 2025
Published online: 27 August 2025

References

1. Appelbaum LG, Erickson G. Sports vision training: a review of the state-of-the-art in digital training techniques. Int Rev Sport Exerc Psychol. 2018;11(1):160–89.
2. Davis J, Bransen L, Devos L, Jaspers A, Meert W, Robberechts P, et al. Methodology and evaluation in sports analytics: challenges, approaches, and lessons learned. Mach Learn. 2024;113(9):6977–7010.
3. Exel J, Dabnichki P. Precision sports science: what is next for data analytics for athlete performance and well-being optimization? Appl Sci. 2024;14(8):3361.
4. Sarlis V, Tjortjis C. Sports analytics—evaluation of basketball players and team performance. Inform Syst. 2020;93:101562.
5. Yu T, Xu Y, Zhang Z, Sun Y, Zhong J, Ding C. The impact of core training on overall athletic performance in different sports: a comprehensive meta-analysis. BMC Sports Sci, Med Rehabil 2025;17(1):112.
6. Seçkin AÇ, Ateş B, Seçkin M. Review on wearable technology in sports: concepts, challenges and opportunities. Appl Sci. 2023;13(18):10399.
7. Torres-Ronda L, Beanland E, Whitehead S, Sweeting A, Clubb J. Tracking systems in team sports: a narrative review of applications of the data and sport specific analysis. Sports Medicine-Open. 2022;8(1):15.
8. Cossich VR, Carlgren D, Holash RJ, Katz L. Technological breakthroughs in sport: current practice and future potential of artificial intelligence, virtual reality, augmented reality, and modern data visualization in performance analysis. Appl Sci. 2023;13(23):12965.
9. Arzehgar A, Seyedhasani SN, Ahmadi FB, Bagheri Baravati F, Sadeghi Hesar A, Kachooei AR, et al. Sensor-based technologies for motion analysis in sports injuries: a scoping review. BMC Sports Sci Med Rehabilitation. 2025;17(1):15.
10. Veiga JJD, O'Reilly M, Whelan D, Caulfield B, Ward TE. Feature-free activity classification of inertial sensor data with machine vision techniques: method, development, and evaluation. JMIR mHealth uHealth. 2017;5(8):e7521.
11. Cronin NJ. Using deep neural networks for kinematic analysis: challenges and opportunities. J Biomech. 2021;123:110460.
12. Pillalamarri R, Shanmugam U. A review on EEG-based multimodal learning for emotion recognition. Artif Intell Rev. 2025;58(5):131.
13. Rahimian P, Toka L. Optical tracking in team sports: a survey on player and ball tracking methods in soccer and other team sports. J Quant Anal Sports. 2022;18(1):35–57.
14. Baumeister R, Leary M. Writing narrative literature reviews, 1. Rev Gen Psychol Https://doi Org. 1997;3(3):311–20. http://doi.org/101037/1089–26801.
15. Siddaway AP, Wood AM, Hedges LV. How to do a systematic review: a best practice guide for conducting and reporting narrative reviews, meta-analyses, and meta-syntheses. Ann Rev Psychol. 2019;70(1):747–70.
16. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372.
17. PRISMAStatement. PRISMA for Searching (PRISMA-Search). 2025 Available from: https://www.prisma-statement.org/prisma-search.
18. Rethlefsen ML, Kirtley S, Waffenschmidt S, Ayala AP, Moher D, Page MJ, et al. PRISMA-S: an extension to the PRISMA statement for reporting literature searches in systematic reviews. Syst Reviews. 2021;10:1–19.
19. Chen L, Wang W. Analysis of technical features in basketball video based on deep learning algorithm. Sig Process Image Commun. 2020;83:115786.
20. Goka R, Moroto Y, Maeda K, Ogawa T, Haseyama M. Multimodal shot prediction based on Spatial-Temporal interaction between players in soccer videos. Appl Sci. 2024;14(11):4847.
21. Kamble PR, Keskar AG, Bhurchandi KM. A deep learning ball tracking system in soccer videos. Opto-Electron Rev. 2019;27(1):58–69.
22. Khobdeh SB, Yamaghani MR, Sareshkeh SK. Basketball action recognition based on the combination of YOLO and a deep fuzzy LSTM network. J Supercomputing. 2024;80(3):3528–53.
23. Li X, Luo R, Islam FU. Tracking and detection of basketball movements using multi-feature data fusion and hybrid YOLO-T2LSTM network. Soft Comput. 2024;28(2):1653–67.
24. Li X, Ullah R. An image classification algorithm for football players' activities using deep neural network. Soft Comput. 2023;27(24):19317–37.
25. Liu W, Yan CC, Liu J, Ma H. Deep learning based basketball video analysis for intelligent arena application. Multimedia Tools Appl. 2017;76:24983–5001.
26. Rahman MA. A deep learning framework for football match prediction. SN Appl Sci. 2020;2(2):165.
27. Thulasya Naik B, Hashmi MF, Gupta A. EloU-distance loss: an automated team-wise player detection and tracking with Jersey colour recognition in soccer. Connection Sci. 2024;36(1):2291991.
28. Yoon Y, Hwang H, Choi Y, Joo M, Oh H, Park I, et al. Analyzing basketball movements and pass relationships using realtime object tracking techniques based on deep learning. IEEE Access. 2019;7:56564–76.
29. Liang C, Liang Z. The application of deep Convolution neural network in volleyball video behavior recognition. IEEE Access. 2022;10:125908–19.
30. Wang H, Jin X, Zhang T, Wang J. Convolutional neural network-based recognition method for volleyball movements. Heliyon. 2023;9(8).
31. Ahmad W, Munsif M, Ullah H, Ullah M, Alsuwailem AA, Saudagar AKJ, et al. Optimized deep learning-based cricket activity focused network and medium scale benchmark. Alexandria Eng J. 2023;73:771–9.
32. Sarma MS, Deb K, Dhar PK, Koshiba T. Traditional Bangladeshi sports video classification using deep learning method. Appl Sci. 2021;11(5):2149.
33. Cao Z, Liao T, Song W, Chen Z, Li C. Detecting the shuttlecock for a badminton robot: A YOLO based approach. Expert Syst Appl. 2021;164:113833.
34. Luo J, Hu Y, Davids K, Zhang D, Gouin C, Li X et al. Vision-based movement recognition reveals badminton player footwork using deep learning and binocular positioning. Heliyon. 2022;8(8).
35. Yang W, Jiang M, Fang X, Shi X, Guo Y, Al-qaness MA. A high-precision and efficient method for badminton action detection in sports using you only look once with hourglass network. Eng Appl Artif Intell. 2024;137:109177.
36. Amosa TI, Sebastian P, Izhar LI, Ibrahim O, Ayinla LS, Bahashwan AA, et al. Multi-camera multi-object tracking: A review of current trends and future advances. Neurocomputing. 2023;552:126558.
37. Huang H-W, Yang C-Y, Sun J, Kim P-K, Kim K-J, Lee K, et al. editors. Iterative scale-up expansioniou and deep features association for multi-object tracking in sports. Proceedings of the IEEE/CVF winter conference on applications of computer vision; 2024.
38. Dargan S, Kumar M, Ayyagari MR, Kumar G. A survey of deep learning and its applications: a new paradigm to machine learning. Arch Comput Methods Eng. 2020;27:1071–92.
39. Cust EE, Sweeting AJ, Ball K, Robertson S. Machine and deep learning for sport-specific movement recognition: a systematic review of model development and performance. J Sports Sci. 2019;37(5):568–600.
40. Alzubaidi L, Zhang J, Humaidi AJ, Al-Dujaili A, Duan Y, Al-Shamma O, et al. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. J Big Data. 2021;8:1–74.
41. Ghosh I, Ramasamy Ramamurthy S, Chakma A, Roy N. Sports analytics review: artificial intelligence applications, emerging technologies, and algorithmic perspective. Wiley Interdisciplinary Reviews: Data Min Knowl Discovery. 2023;13(5):e1496.
42. Sadr MM, Khani M, Tootkaleh SM. Predicting athletic injuries with deep learning: evaluating CNNs and RNNs for enhanced performance and safety. Biomed Signal Process Control. 2025;105:107692.
43. Meng B, Liu X, Wang X. Human action recognition based on quaternion spatial-temporal convolutional neural network and LSTM in RGB videos. Multimedia Tools Appl. 2018;77(20):26901–18.
44. Cui Y. An efficient approach to sports rehabilitation and outcome prediction using RNN-LSTM. Mob Networks Appl. 2024:1–16.
45. Mienye ID, Swart TG, Obaido G. Recurrent neural networks: a comprehensive review of architectures, variants, and applications. Information. 2024;15(9):517.
46. Jozefowicz R, Zaremba W, Sutskever I, editors. An empirical exploration of recurrent network architectures. Int Conf Mach Learn. 2015: PMLR.
47. Ahmed SF, Alam MSB, Hassan M, Rozbu MR, Ishtiak T, Rafa N, et al. Deep learning modelling techniques: current progress, applications, advantages, and challenges. Artif Intell Rev. 2023;56(11):13521–617.
48. Zhao Z, Chai W, Hao S, Hu W, Wang G, Cao S et al. A survey of deep learning in sports applications: perception, comprehension, and decision. IEEE Trans Vis Comput Graph. 2025.
49. Tayyab M, Alateyah SA, Alnusayri M, Alatiyyah M, AlHammadi DA, Jalal A, et al. A hybrid approach for sports activity recognition using key body descriptors and hybrid deep learning classifier. Sensors. 2025;25(2):441.
50. Yücebaş SC. A deep learning analysis for the effect of individual player performances on match results. Neural Comput Appl. 2022;34(15):12967–84.
51. Skelton AFGaJ. Evaluating the key metrics of deep learning models 2025 [Available from: https://www.digitalocean.com/community/tutorials/deep-learning-metrics-precision-recall-accuracy.
52. Srivastava T. 12 Important model evaluation metrics for machine learning everyone should know (Updated 2025) 2025 [Available from: https://www.analyticsvidhya.com/blog/2019/08/11-important-model-evaluation-error-metrics/.

53. Naik BT, Hashmi MF, Bokde ND. A comprehensive review of computer vision in sports: open issues, future trends and research directions. Appl Sci. 2022;12(9):4429.
54. Zhi-chao C, Zhang L. Key pose recognition toward sports scene using deeply-learned model. J Vis Commun Image Represent. 2019;63:102571.
55. Host K, Ivašić-Kos M. An overview of human action recognition in sports based on computer vision. Heliyon. 2022;8(6).
56. He D, Li L. A novel deep learning method based on modified recurrent neural network for sports posture recognition. J Appl Sci Eng. 2021;24(1):43–8.
57. Ren H. Sports video athlete detection based on deep learning. Neural Comput Appl. 2023;35(6):4201–10.
58. Huang R, Zhang B, Yao Z, Xie B, Guo J, DESNet. Real-time human pose Estimation for sports applications combining IoT and deep learning. Alexandria Eng J. 2025;112:293–306.
59. Xiao Y, Tian Z, Yu J, Zhang Y, Liu S, Du S, et al. A review of object detection based on deep learning. Multimedia Tools Appl. 2020;79:23729–91.
60. Hu M, Zhang M, Yu K. Design of sports training information analysis system based on a multi-target visual model under sensor-scale Spatial transformation. PeerJ Comput Sci. 2024;10:e2030.
61. Rafiq M, Rafiq G, Agyeman R, Choi GS, Jin S-I. Scene classification for sports video summarization using transfer learning. Sensors. 2020;20(6):1702.
62. Gao T, Zhang M, Zhu Y, Zhang Y, Pang X, Ying J, et al. Sports video classification method based on improved deep learning. Appl Sci. 2024;14(2):948.
63. Ji S, Xu W, Yang M, Yu K. 3D convolutional neural networks for human action recognition. IEEE Trans Pattern Anal Mach Intell. 2012;35(1):221–31.
64. Aghamohammadi A, Beheshti Shirazi SA, Banihashem SY, Shishechi S, Ranjbarzadeh R, Jafarzadeh Ghoushchi S, et al. A deep learning model for ergonomics risk assessment and sports and health monitoring in self-occluded images. SIViP. 2024;18(2):1161–73.
65. El-Kassabi HT, Khalil K, Serhani MA, editors. Deep learning approach for forecasting athletes' performance in sports tournaments. Proceedings of the 13th International Conference on Intelligent Systems: Theories and Applications; 2020.
66. Musat CL, Mereuta C, Nechita A, Tutunaru D, Voipan AE, Voipan D, et al. Diagnostic applications of AI in sports: a comprehensive review of injury risk prediction methods. Diagnostics. 2024;14(22):2516.
67. Zhang J, Han D, Han S, Li H, Lam W-K, Zhang M, ChatMatch. Exploring the potential of hybrid vision–language deep learning approach for the intelligent analysis and inference of racket sports. Comput Speech Lang. 2025;89:101694.
68. Reis FJ, Alaiti RK, Vallio CS, Hespanhol L. Artificial intelligence and machine-learning approaches in sports: concepts, applications, challenges, and future perspectives. Braz J Phys Ther. 2024:101083.
69. Pietraszewski P, Terbalyan A, Roczniok R, Maszczyk A, Ornowski K, Manilewska D, et al. The role of artificial intelligence in sports analytics: a systematic review and meta-analysis of performance trends. Appl Sci. 2025;15(13):7254.

Publisher's note