Using AI and Computer Vision for Comparative Analysis of Human Actions in Sports Videos

1st Nidhi G. Gupta
Department of Computer Science & Engineering
Sipna College of Engineering & Technology
Amravati, India
nggupta@sipnaengg.ac.in

2nd Dr.A. A. Bardekar
Department of Computer Science & Engineering
Sipna College of Engineering & Technology
Amravati, India
aabardekar@gmail.com

Abstract **The rapid advancement of Artificial Intelligence (AI) and Computer Vision (CV) has transformed the analysis of human performance in sports. This study proposes an automated framework for the comparative analysis of human actions in sports videos using deep learning-based pose estimation and motion recognition techniques. The system leverages convolutional neural networks (CNNs) and recurrent neural networks (RNNs), particularly Long Short-Term Memory (LSTM) models, to extract spatiotemporal features from video sequences. By detecting and tracking key body joints, the model generates pose-based representations that enable precise comparison of athletes' movements across different performances or individuals.**

The proposed method allows for quantitative evaluation of action similarity, technique deviations, and motion efficiency, which can assist coaches and analysts in performance assessment, skill improvement, and injury prevention. Experimental results on benchmark sports datasets demonstrate high accuracy in recognizing and comparing complex motion patterns such as swinging, kicking, and jumping. The framework's modular design also supports real-time analysis, making it suitable for both professional training environments and broadcast analytics.

This research highlights the potential of integrating AI-driven computer vision techniques with sports science to create intelligent systems that offer objective, data-driven insights into athletic performance.

Keywords— Artificial Intelligence, Computer Vision, Deep learning, Human Behavior Recognition, Human Actions. Machine Learning.

I. Introduction

Numerous applications have emerged from the development of this technology, encompassing various aspects of sports production, analytics, and fan engagement. Key capabilities include player tracking and position estimation, where AI models identify and follow athletes' movements throughout a game, and ball trajectory extraction, which provides quantitative data for evaluating accuracy, speed, and tactical performance. Additionally, content extraction and indexing allow large video repositories to be organized and searched efficiently, while automatic summarization and highlight detection enable the generation of concise clips that capture the most significant moments of a match.

Beyond analysis, advanced techniques such as on-demand 3D reconstruction, animation generation, and virtual view synthesis support immersive visualizations that enhance both training and broadcast experiences. These can be further integrated with editorial content creation and virtual content insertion, such as augmented advertisements or virtual replays. Content visualization and enhancement tools improve the presentation of tactical data and facilitate audience understanding.

Moreover, gameplay analysis and evaluation systems allow coaches and analysts to assess team strategies, player actions, and referee decisions with objective metrics. This not only contributes to improved training efficiency but also supports fairer officiating and deeper tactical insights. As a result, automatic video analysis represents a transformative approach that bridges the gap between traditional sports analytics and modern AI-driven performance assessment.[1-2].

The concept of modeling and predicting human behavior and decision-making has been studied for decades across multiple disciplines, including psychology, neuroscience, and computer science. In computer vision, HAR builds upon these foundations by using visual data to model physical movements and behavioral patterns. Traditional approaches relied heavily on handcrafted features such as optical flow, silhouettes, or motion trajectories. However, with the rise of deep learning, modern HAR techniques now utilize powerful architectures like Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Transformers to automatically learn spatiotemporal features from large-scale video datasets.

Understanding human thought and behavior is fundamental to recognizing complex social interactions, gestures, and coordinated group activities. This ability supports a wide range of real-world applications, including surveillance and security systems, sports analytics, healthcare and rehabilitation, human-computer interaction (HCI), autonomous driving, and virtual or augmented reality environments[3-4]. Human social behavior is inherently complex and hierarchical, comprising multiple layers of interaction, coordination, and communication. These structures become increasingly intricate as they evolve, encompassing a wide range of cognitive, emotional, and physical attributes. Despite the inherent complexity and limitations of such social systems, humans have demonstrated exceptional ability in navigating and interpreting these hierarchical structures, effectively understanding both individual and collective behaviors.

Over time, numerous theoretical frameworks have been proposed to explain various aspects of human perception, action understanding, and behavioral prediction. These include models derived from cognitive science, neuroscience, and computer vision, all of which aim to capture the mechanisms underlying human motion and intent recognition. However, in the context of Human Action Recognition (HAR), this study focuses on the computational and visual understanding of these behaviors—specifically, how machines can interpret human actions from image or video sequences using artificial intelligence techniques[5]. The process of Human Action Recognition (HAR) is inherently intricate, as it requires multiple levels of analysis and understanding from visual data. At its core, the task involves identifying a person within an image or video, tracking their movements across spatial and temporal dimensions, and recognizing the specific action being performed. Each of these components—detection, localization, and classification—presents its own set of challenges, especially in dynamic and cluttered environments where multiple individuals may be interacting simultaneously or where occlusions and camera motion occur.

In a typical HAR pipeline, the system must first detect and isolate the human subject from the background using object detection or segmentation techniques. Next, it must analyze motion patterns across consecutive frames to determine the spatiotemporal boundaries of the action—that is, where and when the action begins and ends within the video sequence. Finally, the system must classify the type of action being executed, such as running, throwing, jumping, or waving. Advanced systems may also incorporate pose estimation to extract skeletal keypoints or joint coordinates, allowing for a more precise representation of human motion dynamics[6].

Given the large range of possible applications and data types, there is no one-size-fits-all solution to HAR difficulties. Also, there are a lot of other fields that could have potential uses. Some examples of possible applications include monitoring patient recovery, detecting suspicious activity during CCTV surveillance, analysing player behaviour for different types of human-computer interactions, and so forth. A combination of factors, including the meteoric rise in processing power and speed and the constant development of image processing methods, led to the emergence of computer vision technology.

Fig. 1. Human action recognition (HAR) is part of different Computer Science fields [4]

The primary purpose of Computer Vision (CV) technology is to enable machines to automatically analyze, interpret, and understand visual information from images and videos in a manner similar to human vision. This is achieved through the integration of Image Processing, Machine Learning (ML), and Deep Learning (DL) techniques, which work together to extract meaningful patterns and features from raw visual data. The goal is not only to detect and identify objects or scenes but also to derive high-level semantic information—such as actions, emotions, or environmental context—that supports intelligent decision-making.

In practical applications, computer vision algorithms can be implemented on embedded systems, edge devices, or high-performance computing platforms depending on the complexity and real-time requirements of the task. On embedded platforms, optimized lightweight models allow for efficient execution in constrained environments such as mobile devices, surveillance cameras, drones, and autonomous vehicles. In contrast, more computationally intensive systems deployed on powerful computers or cloud infrastructures can handle large-scale video analytics, 3D reconstruction, and high-resolution image processing[7-8]. The number ten. At the moment, it's the focal point and the area with the most rapid growth. A well-liked subfield of computer vision, video-based human behaviour recognition aims to automatically recognize human behaviour in video or picture sequences by studying and understanding human behaviour in the video, including individual behaviour, the interaction between people, and the interaction between people and the environment [9]. There are several possible applications for video-based human behaviour, including intelligent surveillance, video search, behaviour analysis, and human-computer interaction. Among the many potential applications are those about healthcare, virtual reality, security, intelligence surveillance systems, healthcare, and human-computer interfaces. references [10-11].

The paper is organized as follows: Section II introduced the background related to analysis approach. In section III, statistics of studies in sports are presented. Section IV discuss the various existing research articles and framework for HAR used in sport domain introduced in section V. Section VI discuss comparative analysis of existing research articles. Lastly, we conclude the paper in Section VII.

II. Background

A. *Artificial Intelligence*

One of the most prominent subfields in computer science, Artificial Intelligence (AI) seeks to develop computer systems capable of carrying out activities traditionally associated with humans. The goal is to create computers that can mimic human behaviour, cognition, and decision-making abilities [1]. Computer Vision (CV) is a subfield of AI that focuses on picture and video processing similar to human vision, to comprehend their content and extract valuable information from them [2]. In other words, we want to teach computers human-level intelligence and give them agency [4].

B. *Computer Vision*

Traditional manual analysis methods are no longer feasible due to the sheer scale and velocity of data production. As a result, AI-powered CV systems have become essential for processing and understanding visual content automatically. These systems employ advanced techniques in deep learning, pattern recognition, and data mining to extract meaningful information, detect anomalies, and recognize patterns that would otherwise go unnoticed by human observers.[3]. Online photo-sharing platforms like Facebook and Instagram get over 3 billion daily photo uploads. Also, one of the most popular video search engines, YouTube, has hundreds of hours of user-uploaded videos [4]. Several datasets for CV implementations may be constructed using these data. One factor that has contributed to CV's meteoric rise is the proliferation of high-powered, open-source machine-learning algorithms that have been tested and used in a variety of contexts. Less effort is spent on implementations since popular libraries such as Tensorflow and OpenCV, as well as Facebook and Microsoft libraries, make it unnecessary to start CV projects from scratch. Some of CV's primary functions include validating the presence of an item or activity in the input picture or video, classifying it according to its location, and segmenting it according to its pixels [9].

III. Statistics of studies in sports

The fundamental stage in tracking a player involves accurately determining their position on the field or court at any given time, as this serves as the foundation for further analysis and performance evaluation. This positional data is crucial not only for performance tracking but also for sports graphics systems that analyze and visually capture significant game moments for broadcast and review. Modern commercial broadcast analysis systems employ a range of tools and methods to achieve this objective. Some systems are fully automatic, utilizing advanced computer vision techniques such as image segmentation and motion detection to identify regions in video frames that most likely correspond to players. These systems can process multiple camera feeds simultaneously, ensuring continuous and automated tracking. On the other hand, manual or semi-automatic systems rely on human operators who identify player positions, often by clicking on the players' feet or other reference points using a calibrated camera image that is geometrically aligned with the playing surface. Both methods aim to generate precise location data that can be used for tactical, technical, and performance evaluation. This information is highly valuable to coaches and analysts, as it allows them to study player movements, team formations, spacing, and overall game dynamics in real time. By interpreting this data, coaches can make informed decisions to enhance strategy, adjust formations, and ultimately improve team performance. Player tracking has therefore become an essential component in sports such as badminton, volleyball, hockey, soccer, and many others, enabling teams to gain deeper insights into their playstyle and optimize their competitive advantage through data-driven analysis.

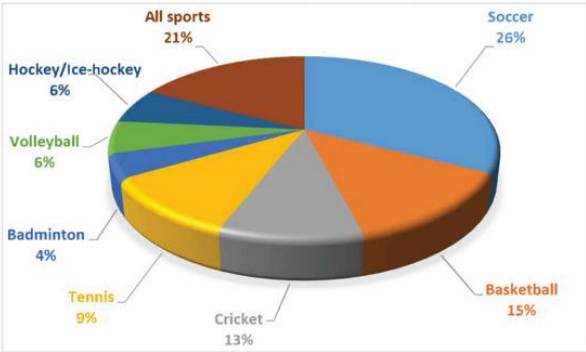

Fig. 2. Sports wise research progress [1]

The research articles are chosen through a peer review process after several articles from different sources are obtained. These sources range from highly influential websites in the fields of player, ball, and referee tracking and detection in sports to object classification in sports, player behaviour, and performance analysis, referee/umpire gesture recognition, automatic highlight detection, and score updating, among other topics. The data from studies of different sports in different applications—such as monitoring players and balls, predicting trajectory, classifying players, and summarizing videos—are shown in Figure 2. All this research has been published in several reputable publications.

IV. Related Work

Early CNN-based approaches primarily relied on 2D CNNs, which process video frames as independent images. While effective for extracting spatial features such as posture, appearance, and object context, these models lack the capability to capture temporal dynamics—the changes in movement over time that are critical for accurately distinguishing between similar-looking actions. To address this limitation, researchers introduced 3D Convolutional Neural Networks (3D CNNs), which extend the conventional 2D convolutional operation by adding a temporal dimension to the filters. This allows the model to simultaneously learn spatiotemporal features, capturing both the appearance and motion information across consecutive video frames.[12]. Research in human organizational behaviour and predictive analytics aims to review the theoretical and empirical frameworks used in these areas throughout the last 20 years [13]. Through extensive training and evaluation on a benchmark dataset of human actions, the proposed model achieved a detection accuracy rate of 92.51%, demonstrating its robustness and reliability in real-world scenarios. This high level of performance indicates the system's ability to correctly classify both the type of action and its duration, enabling it to recognize behaviors that deviate from typical motion patterns, such as running in restricted areas, physical confrontations, or abrupt directional changes. [14]. They look into human behaviour and the links between things like physical characteristics, food choices, and internet use [15]. A set of rules for conduct that has been refined via trial and error with the use of quantitative models of human behaviour [16]. Here we take a look at the present status of action recognition using deep learning techniques and video analysis [17]. An mAP of 79.33% on the frame-based and 84.4% on the image-based assessment of the HAR datasets was reported by several HARS for various human activities seen in the films that were studied and analysed [18]. The feature-contribution-based fusion method considerably enhances performance on both benchmark datasets, with improvements of more than 2% overall and 2.69 percentage points on HMDB51 in particular [19]. In practice, the system processes sequences of frames captured by home cameras or embedded sensors. The 2D CNN extracts relevant features—such as posture, gestures, and object interactions—at each frame and then aggregates this information to understand the context of the activity. This allows the system to detect behaviors that deviate from normal routines, such as unauthorized entry, potential accidents, or abnormal interactions with household objects. [21].

New categories based on Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN) methods are suggested, with a quick comparison of the original feature-based method and its combination with deep learning [22]. Using the widely used deep learning method, a model has been developed to detect abnormal behaviour in movies including violent learning and fainting [23]. An advanced model for interpreting human behaviour in videos using a tensor-train approach was presented [24]. Using data from five separate behavioural tasks in both mice and humans, methods based on deep learning were described for analysing behavioural imaging data [25]. The methods utilized three different convolutional neural network architectures. To efficiently and intelligently analyse the massive social media datasets, an architecture was proposed based on a big data analytics technique [26]. A suggested system for improper human behaviour identification in mental patient monitoring used an unsupervised learning technique based on the N-cut algorithm and Support Vector Machines (SVMs) to label video segments and categorize occurrences into normal and abnormal categories [27]. In a multi-view context, deep networks trained using deep learning techniques like Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks can recognize human activity in films [28]. Support is given to the elderly by keeping tabs on their movements in various indoor and outdoor settings. This is achieved by analysing data acquired from a smartphone's gyroscope and accelerometer, and then applying machine learning and deep learning techniques to identify human activity [29]. A convolutional neural network–based approach to participant behaviour recognition in a video session [30]. By learning boundless transformations informed by flat RGB videos, a new standard for human behaviour understanding with UAVs called UAV-Human has been suggested [31]. This standard provides a fisheye-based action identification system that fixes the distortions seen in fisheye movies.

A summary of approaches and methods based on machine learning for detecting patterns of behaviour in a sequence of images was provided [32]. Learning was achieved with the assistance of the following algorithms: the gate control algorithm, the LSTM long short-term extreme value network algorithm, the RNN recurrent neural network algorithm, the CNN convolutional neural network algorithm dimension, and the HATP algorithm, which categorized samples into 12 groups according to single and composite features [33]. To simulate human actions on undirected and node-attributed networks, a Socially Restricted Boltzmann Machine (SRBM) was suggested as a deep learning model [34]. To identify outliers in video footage, a new approach was devised that employed a computational technique that was derived from a hybrid of the differential evolution algorithm and a cellular neural network [35]. With a focus on video surveillance applications, the described contextual anomalous human behaviour identification is given [36].

V. Framework for HAR in Sport Domain

Human Activity Recognition (HAR) in sports refers to the process of identifying, tracking, and classifying the physical movements and activities of athletes by analyzing data obtained from visual inputs (e.g., videos, motion capture) or sensor-based systems (e.g., wearable devices, inertial measurement units). The primary goal of HAR in sports is to transform raw data into meaningful insights about an athlete's performance, technique, and overall physical behavior, enabling more informed training, coaching, and injury prevention strategies.

The rapid advancement of wearable technology, such as accelerometers, gyroscopes, and smart clothing, has provided continuous and precise measurements of kinematic and biomechanical parameters during athletic activities. Similarly, video analytics powered by computer vision techniques allow for detailed tracking of player movements, joint positions, and interactions in both individual and team sports. By integrating these data sources with machine learning algorithms, HAR systems can automatically detect patterns, classify complex actions, and identify deviations from optimal performance.

In sports, the term "recognition of human action" refers to the use of computer vision techniques to identify players' or athletes' movements. The actions of athletes or players can be observed during warm-ups, fitness sessions, sport-specific training, games, or contests. As a result, HAR may be used to track athletes' performance, evaluate different players' actions, or a single player's repeated repetitions of a technique to aid with technique practice, style improvement, and other purposes. Furthermore, automatic statistical analysis of a sporting event or the performances of certain athletes may be given using HAR. Figure 3 illustrates a common supervised learning job using sports data called HAR in sports.

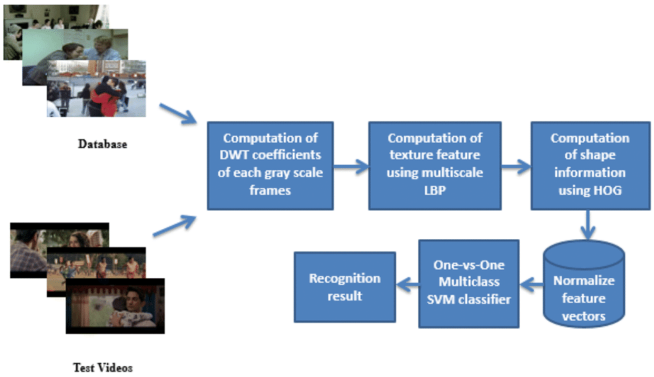

Fig. 3. The workflow of the process of Human action recognition [4]

The first steps involve gathering and labeling data for a particular job, including preprocessing operations like feature extraction and digital noise removal. Deep learning (DL) made feature extraction automatic, replacing the laborious process of standard machine learning (ML) approaches. Hand-crafted feature-based methods consist of two basic steps: feature extraction and, if necessary, dimensionality reduction before classification. Feature extraction may be thought of as a first step in data pre-processing that eliminates superfluous information. Low-level and high-level features are the two categories of features. Corners, edges, and curves are the essential components of low-level features. To obtain organized information about the activity being conducted, high-level features should incorporate domain knowledge. A variety of characteristics may be employed for HAR in sports, including Histogram of Oriented Gradients (HOG) and Optical Flow for extracting motion information. Although the HAR feature extraction process could make it possible to determine how objects move in the picture, these descriptors cannot explain the actions. As a result, machine learning approaches for categorization are often employed. Other DL-based methods for classifying actions in HAR include automated feature extraction from photos, description, and classification.

Once features have been extracted, the next step is training machine learning or deep learning models for classifying activities. After model training, the HAR system can classify the athlete's activities in real-time or from recorded data. After the activities have been recognized, post-processing can be done to derive insights and visualizations that are useful for performance analysis or injury prevention. The ultimate goal of HAR in sports is to provide feedback to athletes and coaches. The system can deliver real-time insights or historical reports. The final phase is the deployment of the HAR system in real-world environments, such as during live sports events, training sessions, or rehabilitation.

Implementing HAR in the sports domain offers numerous benefits for athletes, coaches, and teams. By integrating sensor data, video analytics, and machine learning, this framework enables effective monitoring, feedback, and improvement in sports performance. Real-time recognition of activities can lead to better-informed coaching decisions and help athletes avoid injuries, optimize performance, and achieve their full potential.

VI. Comparative Analysis

Table 1 depicts the summary of existing research articles related to human behavioral while analyzing the action in sport videos. It also presented methodology used with dataset used in their implementation considering the context of applications. Also, figure 4 depicts the accuracy obtained by various researchers in their work.

TABLE I. SUMMARY OF EXISTING RESEARCH ARTICLES

Ref	Input Consideration	Datasets	Intelligent Classifiers	Results Evaluation
P K Sarkar et.al. [14]	Human Activity Sequences	Real Time	CNN	Precision = 87%, Recall = 99%, F-score = 93%
Xiwei Liu et.al. [19]	Human Behaviour in Sports	UCF101 and HMDB51	R3D model	Accuracy = 87.89%
Alaeddine Mihoub et.al. [20]	Human Activity in Smart Homes	Orange4 Home	RNN	Accuracy = 94.09%, F-score = 94.26%
J Indhumathi et.al. [21]	Human Suspicious Activity	Kaggle and Real time	2D-CNN	Accuracy = 99.18%, Precision = 99.03%, Recall = 99.02%, F-score = 99.01%.
Ahmet Arac et.al. [25]	Human Motor Behavior in clinic	Real Time	DL	Accuracy = 95.50%
Shih-Chung Hsu et.al. [27]	Human Behavior Detection for Psychiatric Patient Monitoring	Real Time	CRF	Accuracy = 92.4%
Yu-Ling Hsueh et.al. [28]	Human Behavior Recognition from Multiview Videos	MNIST and COIL-20	Stacked Convolutional Autoencoder (SCAE)	Accuracy = 95%, Precision = 98%, Recall = 98%, F-score = 93%
Ahatsham Hayat et.al. [29]	Human Activity Recognition for Elderly People	UCI	Random Forest, SVM, KNN, ANN, LSTM	Accuracy = 95.04%, Precision = 92.87%, Recall = 85.32%, F-score = 88.94%.
Yizhen Meng	Human Behavior Recognition in Video Session	UCF-101, HMDB-51	VGG13-CNN	Accuracy = 96.1%

Ref	Input Consideration	Datasets	Intelligent Classifiers	Results Evaluation
et.al. [30]				
Tianjiao Li et.al. [31]	Human Behavior Understanding with Unmanned Aerial Vehicles	UAV-Human	Guided Transformer I3D (GT-I3D)	Accuracy = 89.6%, mAP = 85.71%
Nhathai Phan et.al. [34]	Human Behavior Prediction in Health Social Networks	Real world	Social Restricted Boltzmann machine (SRBM)	Accuracy = 91.5%
BAI Ya-meng et.al. [35]	Abnormal Human Behavior in Video Images	UCSD Anomaly Detection	ANN	Accuracy = 97.2%, Sensitivity = 98.6%

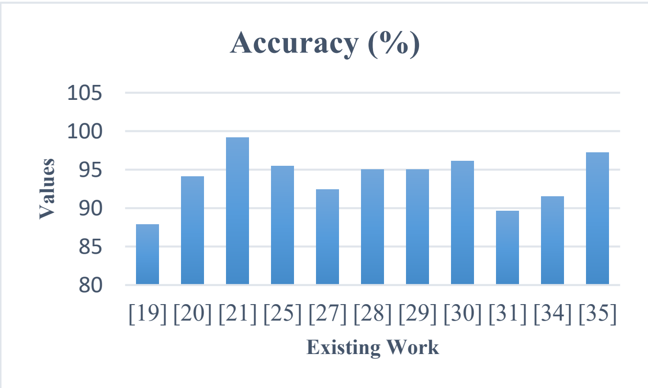

Fig. 4. Accuracy of existing research work.

VII. Conclusion

Sports video analysis has emerged as a rapidly developing and dynamic field of research that integrates advanced technologies from artificial intelligence (AI) and computer vision to transform the way sports performance is analyzed and understood. This paper emphasizes the transformative potential of these technologies in reshaping sports analytics by automating tasks that were once entirely manual and subjective. Through the application of AI-based methods, particularly deep learning and image processing, it becomes possible to detect, classify, and interpret various elements within sports videos—such as player movements, team formations, and game strategies—with high accuracy and in real time.

To ensure the applicability of these methods across different sports, the study presents a comprehensive review of sport-specific, publicly available datasets. Each dataset is analyzed in terms of its content, scope, and suitability for particular computer vision tasks, such as player detection, pose estimation, action recognition, and event classification. Additionally, the paper compares traditional computer vision techniques with modern AI approaches, highlighting their respective advantages, limitations, and performance characteristics. While classical techniques rely heavily on handcrafted features and image segmentation, AI-driven methods can automatically learn complex spatial and temporal patterns from large datasets, offering greater robustness and adaptability to diverse sporting environments.

The study further demonstrates that AI models can be used to classify player positions and roles based on tactical and technical performance metrics derived solely from game data collected across multiple leagues and seasons. By leveraging deep neural networks and video-based motion analysis, these systems are capable of providing detailed insights into player behavior, decision-making, and spatial dynamics throughout a match. This advancement allows coaches and analysts to gain deeper, data-driven insights into team performance, identify areas for improvement, and support strategic decisions related to player selection, training, and in-game tactics.

Moreover, this research highlights the practical benefits of AI-enhanced sports video analysis beyond performance assessment. For instance, club scouting departments can utilize these tools to evaluate how players adapt their playing styles under different conditions or when subjected to various tactical systems. This could be invaluable for predicting a player's potential fit within a team or for understanding how managerial decisions influence player positioning and overall performance.

However, the paper also recognizes several ongoing challenges in this domain. The most significant obstacles include the need for large and diverse annotated datasets to train AI models effectively, the high computational cost associated with deep learning algorithms, and the complexity of generalizing models across different sports, camera setups, and environmental conditions. Additionally, ensuring real-time processing capabilities remains a critical concern for live match applications.

Looking ahead, future research is expected to focus on improving model precision, enhancing the efficiency of real-time analysis, and addressing domain-specific issues such as occlusion, motion blur, and player identification. These advancements will contribute not only to more accurate and automated sports video analysis systems but also to a richer and more nuanced understanding of human behavior and decision-making in competitive sports. Ultimately, this evolving field promises to revolutionize the way sports are analyzed, coached, and experienced, bridging the gap between data science and athletic performance through the power of AI-driven insights.

References

[1] B. T. Naik, M. F. Hashmi and N. D. Bokde, "A Comprehensive Review of Computer Vision in Sports: Open Issues, Future Trends and Research Directions," Applied Sciences 12, no. 9: 4429, 2022. https://doi.org/10.3390/app12094429.

[2] Z. Zhao, W. Chai, S. Hao, W. Hu, G. Wang, S. Cao, M. Song, J. Hwang, G. Wang, "A Survey of Deep Learning in Sports Applications: Perception, Comprehension, and Decision," Computer Vision and Pattern Recognition, 2023. https://doi.org/10.48550/arXiv.2307.03353.

[3] B. Ma, S. Nie, M. Ji, J. Song, and W. Wang, "Research and Analysis of Sports Training Real-Time Monitoring System Based on Mobile Artificial Intelligence Terminal," Wireless Communications and Mobile Computing, vol. 2020, pp. 1–10, Sep. 2020, doi: https://doi.org/10.1155/2020/8879616.

[4] K. Host and M. Ivašić-Kos, "An overview of Human Action Recognition in sports based on Computer Vision," Heliyon, vol. 8, no. 6, p. e09633, Jun. 2022, doi: https://doi.org/10.1016/j.heliyon.2022.e09633.

[5] M. Chen and Y. Zhou, "Analysis of Students' Sports Exercise Behavior and Health Education Strategy Using Visual Perception–Motion Recognition Algorithm," Frontiers in Psychology, vol. 13, May 2022, doi: https://doi.org/10.3389/fpsyg.2022.829432.

[6] J. G. Claudino, D. de O. Capanema, T. V. de Souza, J. C. Serrão, A. C. Machado Pereira, and G. P. Nassis, "Current Approaches to the Use of Artificial Intelligence for Injury Risk Assessment and Performance Prediction in Team Sports: a

Systematic Review," Sports Medicine - Open, vol. 5, no. 1, Jul. 2019, doi: https://doi.org/10.1186/s40798-019-0202-3.

[7] A. C. N. Rodrigues, A. S. Pereira, R. M. S. Mendes, A. G. Araújo, M. S. Couceiro, A. J. Figueiredo, "Using Artificial Intelligence for Pattern Recognition in a Sports Context" Sensors 2020, no. 11: 3040. https://doi.org/10.3390/s20113040.

[8] J. Bridgeman and A. Giráldez-Hayes, "Using artificial intelligence-enhanced video feedback for reflective practice in coach development: benefits and potential drawbacks," Coaching: An International Journal of Theory, Research and Practice, pp. 1–18, Jul. 2023, doi: https://doi.org/10.1080/17521882.2023.2228416.

[9] A. Efe, "An Assessment Over the Impact of Artificial Intelligence on Sports Activities and the Sports Industry", 2023.

[10] A. García-Aliaga, M. Marquina, J. Coterón, A. Rodríguez-González, and S. Luengo-Sánchez, "In-game behaviour analysis of football players using machine learning techniques based on player statistics," International Journal of Sports Science & Coaching, vol. 16, no. 1, p. 174795412095976, Sep. 2020, doi: https://doi.org/10.1177/1747954120959762.

[11] M. Freiberg et al., "Machine Learning Classifications of Player Behavior," SSRN Electronic Journal, 2021, doi: https://doi.org/10.2139/ssrn.3824485.

[12] P. Kuppusamy and V. C. Bharathi, "Human abnormal behavior detection using CNNs in crowded and uncrowded surveillance – A survey," Measurement: Sensors, vol. 24, p. 100510, Dec. 2022, doi: https://doi.org/10.1016/j.measen.2022.100510.

[13] R. Oliveira and P. Rodrigues, "A Review of Literature on Human Behaviour and Artificial Intelligence: Contributions Towards Knowledge Management," The Electronic Journal of Knowledge Management, 2021. 19(2), pp. 165-179

[14] P. Sarkar, A. Guho, and A. B. M. Abdullah, "Diagnosing Suspects by Analyzing Human Behavior to Prevent Crime by Using Deep and Machine Learning," SSRN Electronic Journal, 2022, doi: https://doi.org/10.2139/ssrn.4181493.

[15] X. Liu, "Sports Deep Learning Method Based on Cognitive Human Behavior Recognition," Computational Intelligence and Neuroscience, vol. 2022, pp. 1–11, Aug. 2022, doi: https://doi.org/10.1155/2022/2913507.

[16] A. Mihoub, "A Deep Learning-Based Framework for Human Activity Recognition in Smart Homes," Mobile Information Systems, vol. 2021, pp. 1–11, Sep. 2021, doi: https://doi.org/10.1155/2021/6961343.

[17] Indhumathi. J, Balasubramanian. M, and Balasaigayathri. B, "Real-Time Video based Human Suspicious Activity Recognition with Transfer Learning for Deep Learning," International journal of image, graphics and signal processing, vol. 15, no. 1, pp. 47–62, Feb. 2023, doi: https://doi.org/10.5815/ijigsp.2023.01.05.

[18] V. Sharma, M. Gupta, A. K. Pandey, D. Mishraa and A. Kumar, "A Review of Deep Learning-based Human Activity Recognition on Benchmark Video Datasets," Applied Artificial Intelligence. Ajai. 2022. 36. 10.1080/08839514.2022.2093705.

[19] J. Lee and S.-J. Shin, "A Study of Video-Based Abnormal Behavior Recognition Model Using Deep Learning," International journal of advanced smart convergence , vol. 9, no. 4, p. 115–119, Dec. 2020.

[20] C. Dai, X. Liu, H. Xu, L. T. Yang and M. J. Deen, "Hybrid Deep Model for Human Behavior Understanding on Industrial Internet of Video Things," in IEEE Transactions on Industrial Informatics, vol. 18, no. 10, pp. 7000-7008, Oct. 2022, doi: 10.1109/TII.2021.3058276.

[21] A. Arac, P. Zhao, B. H. Dobkin, S. T. Carmichael, and P. Golshani, "DeepBehavior: A Deep Learning Toolbox for Automated Analysis of Animal and Human Behavior Imaging Data," Frontiers in Systems Neuroscience, vol. 13, May 2019, doi: https://doi.org/10.3389/fnsys.2019.00020.

[22] L. F. Borja-Borja, M. Saval-Calvo & J. Azorin-Lopez, "Machine Learning Methods from Group to Crowd Behaviour Analysis," Advances in Computational Intelligence. IWANN 2017. Lecture Notes in Computer Science, vol 10306. Springer, Cham. https://doi.org/10.1007/978-3-319-59147-6_26

[23] J. Zhu, S. B. Goyal, C. Verma, M. S. Raboaca, and T. C. Mihaltan, "Machine Learning Human Behavior Detection Mechanism Based on Python Architecture," Mathematics, vol. 10, no. 17, p. 3159, Sep. 2022, doi: https://doi.org/10.3390/math10173159.

[24] N. Phan, D. Dou, B. Piniewski, and D. Kil, "A deep learning approach for human behavior prediction with explanations in health social networks: social restricted Boltzmann machine (SRBM+)," Social Network Analysis and Mining, vol. 6, no. 1, Sep. 2016, doi: https://doi.org/10.1007/s13278-016-0379-0.

[25] B. Ya-meng, W. Yang, and W. Shen-shen, "Detection of Abnormal Human Behavior in Video Images based on a Hybrid Approach," International Journal of Advanced Computer Science and Applications, vol. 13, no. 11, 2022, doi: https://doi.org/10.14569/ijacsa.2022.0131138.

[26] O. P. Popoola and Kejun Wang, "Video-Based Abnormal Human Behavior Recognition—A Review," IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 42, no. 6, pp. 865–878, Nov. 2012, doi: https://doi.org/10.1109/tsmcc.2011.2178594.

[27] M. U. Tariq, M. Babar, M. Poulin, A. S. Khattak, M. D. Alshehri, and S. Kaleem, "Human Behavior Analysis Using Intelligent Big Data Analytics," Frontiers in Psychology, vol. 12, Jul. 2021, doi: https://doi.org/10.3389/fpsyg.2021.686610.

[28] S. -C. Hsu, C. -H. Chuang, C. -L. Huang, P. -R. Teng and M. -J. Lin, "A video-based abnormal human behavior detection for psychiatric patient monitoring," 2018 International Workshop on Advanced Image Technology (IWAIT), Chiang Mai, Thailand, 2018, pp. 1-4, doi: 10.1109/IWAIT.2018.8369749.

[29] Y.-L. Hsueh, W.-N. Lie, and G.-Y. Guo, "Human Behavior Recognition from Multiview Videos," Information Sciences, vol. 517, pp. 275–296, May 2020, doi: https://doi.org/10.1016/j.ins.2020.01.002.

[30] A. Hayat, F. Morgado-Dias, B. Bhuyan, and R. Tomar, "Human Activity Recognition for Elderly People Using Machine and Deep Learning Approaches," Information, vol. 13, no. 6, p. 275, May 2022, doi: https://doi.org/10.3390/info13060275.

[31] Y. Meng and J. Zhang, "An effective behavior recognition method in the video session using convolutional neural network," PLOS ONE, vol. 17, no. 8, p. e0266734, Aug. 2022, doi: https://doi.org/10.1371/journal.pone.0266734.

[32] T. Li, J. Liu, W. Zhang, Y. Ni, W. Wang, Z. Li, "UAV-Human: A Large Benchmark for Human Behavior Understanding with Unmanned Aerial Vehicles," Computer Vision and Pattern Recognition, 2021. 10.1109/CVPR46437.2021.01600.

[33] M. G. T.Sajana, "Human Behavior Prediction and Analysis Using Machine Learning-A Review," Turkish Journal of Computer and Mathematics Education (TURCOMAT), vol. 12, no. 5, pp. 870–876, Apr. 2021, doi: https://doi.org/10.17762/turcomat.v12i5.1499.

[34] A. S. Fuchs, A. Passarella and M. Conit, "Modeling Human Behavior Part I -- Learning and Belief Approaches," 2022. http://dx.doi.org/10.48550/arXiv.2205.06485

[35] H. H. Pham, L. Khoudour, A. Crouzil, P. Zegers, S. A. Velastin. "Video-based Human Action Recognition using Deep Learning: A Review," 2022. https://doi.org/10.48550/arXiv.2208.03775.

[36] M. Dhar, B. Chakraborty, "A Video Based Human Detection and Activity Recognition – A Deep Learning Approach," Turkish Journal of Computer and Mathematics Education Vol.11No.01 (2020), 551-559.